

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 18, 2005, 23:59:55 ; Search time 133.5 Seconds

(without alignments)  
3702.470 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163

Sequence: 1 atcgctgcctccgacgcg.....ataaccaggtgaccatcaag 639

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_095900\_16266/app\_query.fasta\_1.775  
-DB=A\_Geneseq\_16Dec04 -QFW=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFW=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10030225 @CGN 1.1 224 @runat\_18022005\_095900\_16266 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_16Dec04:  
1: Geneseqp1980s:  
2: Geneseqp1990s:  
3: Geneseqp2000s:  
4: Geneseqp2001s:  
5: Geneseqp2002s:  
6: Geneseqp2003as:  
7: Geneseqp2003bs:  
8: Geneseqp2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1115	95.9	571	4 AAB49765	Human pro
2	1115	95.9	571	4 AAB88393	Human mem
3	430	37.0	81	4 AAM13752	Peptide #
4	430	37.0	81	4 ABB32685	Peptide #
5	430	37.0	81	4 AAM26150	Peptide #
6	430	37.0	81	4 ABB27526	Human pep
7	430	37.0	81	4 ABB18176	Protein #
8	430	37.0	81	4 AAM65885	Human bon
9	430	37.0	81	4 AAM35507	Human bra
10	430	37.0	81	4 ABG47540	Human liv

11	430	37.0	81	4 AAM01497	Peptide #
12	430	37.0	81	5 ABG35520	Human pep
13	230	19.8	173	7 ADF59409	Human pol
14	154	13.2	19938	6 ABP76679	Streptomy
15	148.5	12.8	546	4 ABUS3252	Human tes
16	148.5	12.8	1270	8 ADK57911	Human ext
17	148.5	12.8	1299	4 AAM24322	Human EST
18	148.5	12.8	1311	8 ADK67912	Human ext
19	148.5	12.8	1320	7 ADK65819	Angiogene
20	148.5	12.8	1404	2 AAR26049	MSF precu
21	148.5	12.8	1404	4 AAB29773	Human meg
22	148.5	12.8	1404	4 AAB60568	Human meg
23	148.5	12.8	1404	7 ADK65839	Angiogene
24	148.5	12.8	1404	8 ADM98014	Human meg
25	148.5	12.8	1415	4 AAU32262	Novel hum
26	148	12.7	513	4 ABUS3254	Human tes
27	140.5	12.1	551	4 ABUS3253	Human tes
28	135.5	11.7	377	5 ABP64957	Human pro
29	135	11.6	783	7 ABO69289	Pseudomon
30	134	11.5	399	3 AAB43375	Human ORF
31	133	11.4	339	7 ABO73785	Pseudomon
32	133	11.4	923	4 ABG08600	Novel hum
33	132	11.3	246	7 ABO71543	Pseudomon
C	130.5	11.1	2655	7 ADO59401	Antheraea
35	130	11.2	489	4 ABG14285	Novel hum
36	129	11.1	1518	6 ABJ18375	Breast sp
37	128.5	11.0	386	7 ABO75475	Pseudomon
38	128	11.0	373	7 ABO80173	Pseudomon
39	128	11.0	410	7 ABO77066	Pseudomon
40	128	11.0	554	4 ABG14613	Novel hum
41	128	11.0	655	8 ABM84305	Human dia
42	127.5	11.0	223	8 ADS30809	Bacterial
43	127.5	11.0	233	7 ABO84024	Pseudomon
44	127.5	11.0	1953	6 AAE36881	Nephila m
45	127	10.9	806	8 ADO17018	Predicted

ALIGNMENTS

RESULT 1

AAB49765

ID AAB49765 standard; protein; 571 AA.

XX

AC AAB49765;

XX

DT 20-APR-2001 (first entry)

XX

DE Human proliferation differentiation factor amino acid sequence.

XX

KW Human; proliferation differentiation factor; haematopoietic function.

XX

OS Homo sapiens.

XX

PN WO200104312-A1.

XX

PD 18-JAN-2001.

XX

PF 06-JUL-2000; 2000WO-JP004514.

XX

PR 08-JUL-1999; 99JP-00194179.

PR

18-OCT-1999; 99US-0159586P.

XX

PA (HELI-) HELIX RES INST.

XX

FI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

XX

WPI; 2001-138354/14.

DR

N-PSDB; AAF29348.

XX

PT Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cDNA library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.

XX













33

PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-476286/51.  
 DR Novel single exon nucleic acid probe used to measuring gene expression in  
 XX a human breast.  
 PT Claim 27; SEQ ID NO 10237; 322pp; English.  
 PS The present invention relates to novel single exon nucleic acid probes  
 XX (see AAI00010-AA110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer; disorders of development,  
 CC inflammatory diseases of the breast; fibrocystic changes; proliferative,  
 CC breast disease and non-carcinoma tumours. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 81 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 1.45e-27 Length: 81  
 Score: 430.00 Matches: 79  
 Percent Similarity: 98.77% Conservative: 1  
 Best Local Similarity: 97.53% Mismatches: 1  
 Query Match: 36.97% Indels: 0  
 DB: 4 Gaps: 0

US-10-030-225-1\_COPY\_58\_696 (1-639) x AAM01497 (1-81)

Qy 142 GTCTCAGCTCCCGAGATCCTAGGCTCTGAAGNAGAGGAGGACCACTGCTCC 201  
 Db 1 ValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeuPro 20  
 Qy 202 AGAACCCACCTGCAGGAGAGCCACCAACATGATGCTGCTGCTGCTGCTGCTG 261  
 Db 21 ArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAla 40  
 Qy 262 GCCATGCCCGGAGGACACACACCCCTCCAGGACCCAGAGTTACTCCGTCGGCTG 321  
 Db 41 AlaMetThrProGlyAsnAlaThrProProArgThrProGluValThrProLeuArgLeu 60  
 Qy 322 GAGTCGAGAGCTCCCGGATTGCCAGCACACCTTGAGTACCCCTAACCTGTATCC 381  
 Db 61 GluLeuGlnLysLeuProGlyLeuAlaAsnThrThrLeuSerThrProAsnProAspThr 80  
 Qy 382 CAG 384  
 Db 81 Gln 81

RESULT 12  
 ABG35520

ID ABG35520 standard; peptide; 81 AA.

AC ABG35520;

XX 19-AUG-2002 (first entry)

DT Human peptide encoded by genome-derived single exon probe SEQ ID 25185.

DE Human; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 XX hyaline membrane disease.  
 XX Homo sapiens.  
 XX WO200186003-A2.  
 XX 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US000665.  
 PF 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 DR Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples.

Claim 27; SEQ ID NO 25185; 634pp; English.

The invention relates to a spatially-addressable set of single exon  
 nucleic acid probes for measuring gene expression in a sample derived  
 from human lung comprising single exon nucleic acid probes having one of  
 12614 nucleic acid sequences mentioned in the specification, or their  
 complements or the 12387 open reading frames derived from the 12614  
 probes. Also included are a microarray comprising the novel set of probes  
 and the novel set of probes which hybridise at high stringency to a nucleic  
 acid expressed in the human lung; measuring gene expression in a sample  
 derived from human lung, comprising (a) contacting the array with a  
 collection of detectably labeled nucleic acids derived from human lung  
 mRNA, and (b) measuring the label detectably bound to each probe of the  
 array; identifying exons in a eukaryotic genome, comprising (a)  
 algorithmically predicting at least one exon from genomic sequences of  
 the eukaryote; and (b) detecting specific hybridisation of detectably  
 labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 having a fragment identical to the predicted exon, the probe is included  
 in the above mentioned microarray; assigning exons to a single gene,  
 comprising (a) identifying exons from genomic sequence by the method  
 above and (b) measuring the expression of each of the exons in several  
 tissues and/or cell types using hybridisation to a single exon  
 microarrays having a probe with the exon, where a common pattern of  
 expression of the exons in the tissues and/or cell types indicates that  
 the exons should be assigned to a single gene; a peptide comprising one  
 of 12011 sequences, mentioned in the specification, or encoded by the  
 probes/open reading frames (ORF). The probes are used for gene expression  
 analysis, and for identifying exons in a gene, particularly using human  
 lung derived mRNA and for the study of lung diseases such as asthma, lung  
 cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 present sequence is a peptide/protein encoded by a single exon probe of  
 the invention. Note: The sequence data for this patent did not form part  
 of the printed specification, but was obtained in electronic format  
 directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences







```
Db      90 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh 110
Qy      271 -----CCAGGCAACACCAACCCCTCCAGGACCC 299
Db      110 rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl 130
Qy      300 AGAGTTACTCCGTTGGCTGGAGCTGCAGAAGCTGCCGGGATGGCCAGCACAACTT 359
Db      130 aProThrThrProLysLysLeuThrProThrProLysLysLeuAlaProThrThrPr 150
Qy      360 GAGTACCCTAACCTGATACCCAGGCTTCAGCCTCCCA-----GATCCTAG 407
Db      150 oLulysProAlaProThrThrProGluGluLeuAlaProThrThrProGluGluProTh 170
Qy      408 GCCTCTGAGGGAAGAGGAGGAGGACGACTGCTCCCAAGAACCCACCTGCAGGCAGAGCT 467
Db      170 rProThrThrProGluGluProAlaProThrThrProLysAlaAlaProAsnThrPr 190
Qy      468 ACACCAACATGGTGTGGACTGTCACTAGCCAGCAGCCCTGACC-----CCAGG 518
Db      190 oLysGluProAlaProThrThrProLysGluProAlaProThrProLysGluProAl 210
Qy      519 GAATGCCACGCTCCAGGACCCAGGAGTTACTCCCTTGCTGGAGCTGCAGAGCT 578
Db      210 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrLe 230
Qy      579 GCCAGAAATTGTCACGCAACCTTGAGTACCCCTAACCCCT 618
Db      230 uLysGluProAlaProThrThrProLysLysProAlapro 243
```

Search completed: February 19, 2005, 00:16:30  
Job time : 152.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 19, 2005, 00:09:26 ; Search time 31.5 Seconds  
(without alignments)  
3028.619 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163

Sequence: 1 atcgctgcgtccgcagcgcg.....ataaccaggtgacatcaag 639

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPRO.spool\_p/US10030225/runat\_18022005\_095901\_16296/app\_query.fasta\_1.775  
-DB=Issued\_Patents\_AA-QMFASTAN-SUFFIX=ra1-MINMATCH=0.1-LOOPCL=0  
-LOOPEXT=0-UNITS=bits-START=1-END=1-MATRIX=blosum62-TRANS-human40.cdi  
-LIST=45-DOCALIGN=200-THR SCORE=ptc-THR MAX=100-THR MIN=0-ALIGN=15  
-MODE=LOCAL-OUTFM=ptc-NORM=ext-HEAPSIZ=500-MINLEN=0-MAXLEN=2000000000  
-USER=US10030225@cgn 1.1.46 @runat\_18022005\_095901\_16296 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148.5	12.8	941	4	US-07-757-022B-14
2	148.5	12.8	1022	4	US-07-757-022B-84
3	148.5	12.8	1038	4	US-07-757-022B-74
4	148.5	12.8	1049	4	US-07-757-022B-58
5	148.5	12.8	1140	4	US-07-757-022B-104
6	148.5	12.8	1270	4	US-07-757-022B-44
7	148.5	12.8	1311	4	US-07-757-022B-42
8	148.5	12.8	1313	4	US-07-757-022B-142
9	148.5	12.8	1314	4	US-07-757-022B-50
10	148.5	12.8	1320	4	US-07-757-022B-46
11	148.5	12.8	1320	4	US-07-757-022B-60
12	148.5	12.8	1320	4	US-10-164-595-58

13	148.5	12.8	1354	4	US-07-757-022B-48
14	148.5	12.8	1361	4	US-07-757-022B-40
15	148.5	12.8	1363	4	US-07-757-022B-52
16	148.5	12.8	1404	4	US-07-757-022B-2
17	148.5	12.8	1404	4	US-07-757-022B-62
18	148.5	12.8	1404	4	US-10-164-595-78
19	148.5	12.8	1404	4	US-09-298-970A-1
20	148.5	12.8	1411	4	US-09-949-016-10827
21	135	11.6	783	4	US-09-252-991A-18035
22	133	11.4	339	4	US-09-252-991A-22531
23	132	11.3	246	4	US-09-252-991A-20289
24	128.5	11.0	386	4	US-09-252-991A-24221
25	128	11.0	373	4	US-09-252-991A-28919
26	128	11.0	410	4	US-09-252-991A-25912
27	127.5	11.0	233	4	US-09-252-991A-32770
28	126.5	10.9	503	3	US-08-999-774A-10
29	126.5	10.9	825	4	US-10-210-428-1
30	126.5	10.9	825	4	US-10-237-551-161
31	126.5	10.9	826	4	US-09-894-998A-47
32	126.5	10.9	826	4	US-10-237-551-47
33	126	10.8	226	4	US-09-252-991A-19845
34	126	10.8	863	4	US-09-252-991A-26099
35	125.5	10.8	542	4	US-09-252-991A-21753
36	125.5	10.8	551	4	US-09-949-016-8970
37	125	10.7	254	4	US-09-252-991A-26135
38	124	10.7	585	4	US-09-252-991A-31529
39	123.5	10.6	183	6	5168049-3
40	123.5	10.6	183	6	5168049-3
41	123.5	10.6	405	4	US-09-252-991A-27573
42	123.5	10.6	538	4	US-09-949-016-11715
43	123.5	10.6	580	4	US-09-616-289-47
44	123.5	10.6	580	4	US-09-949-016-11716
45	123.5	10.6	663	4	US-09-252-991A-30843

#### ALIGNMENTS

#### RESULT 1

US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-14

Alignment Scores:
Pred. No.: 9,19e-05 Length: 941
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
DB: 4 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-14 (1-941)
QY 18 CCGAGCCGGCTCTCTCTGCTGCTGCTGCGCGCTGCTGCGGCGGCTAGG 77
Db 299 ProAlaProThrThrProLysGluProAlaProThrThrLysGluProSerProThr 318
QY 78 GCTCCCGCTGAGAGCC-----CGG 98
Db 319 ThrProLysGluProAlaProThrThrLysSerAlaProThrThrLysGluPro 338
QY 99 GCTCCCGGAGCACCGGCTGGGAGCTCACGAGGCTCGCAGAGGTCTCA-----GCCTC 152
Db 339 AlaProThrThrLysSerAlaPro-ThrThrProLysGluProSerProThrThrTh 358
QY 153 CCCAGATCTAGGCTCTGAAGAGAGAGAGAGGACGACCTGCTCCCGACCAACCCACT 212
Db 358 rlysgluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl 378
QY 213 GCAGGACAGGACACACCACTGGATGCTGACTGCTACTGAGCAGGACGACCATGACC-- 270
Db 378 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh 398
QY 271 -----CCAGGCAACACCCCTCCCGAGCAGCC 299
Db 398 rlysgluProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl 418
QY 300 AGAGGTACTCGGTGCGGCTGAGCTGAGAGTGGCGGATGGCGAGGACCAACCTT 359
Db 418 aProThrThrProLysLysLeuThrProThrThrProLysLysLeuAlaProThrThrPr 438
QY 360 GAGTACCCCTAACCTGATACCCAGGCTCAGCTCCCA-----GATCCTAG 407
Db 438 oGluLysProAlaProThrThrProGluLysLeuAlaProThrThrProGluLysProth 458
QY 408 GCCTCTGAGGGAAGAGGAGGAGGACGACTGCTCCCGAGAACCCACCTGCGAGGAGCT 467
Db 458 rProThrThrProGluLysProAlaProThrThrProLysAlaAlaAlaProAsnThrPr 478
QY 468 ACACCAACATGGATGTGACTGTCTGAGCAGGACGACCCCTGACC-----CCAGG 518
Db 478 oLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAl 498
QY 519 GAATGCCACGCTCCCGAGGACGAGGTACTCCCTGCTGCTGAGCTGAGCTCAGAGCT 578
Db 498 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrLe 518
QY 579 GCCAGATTGGTCCAGCAACCTTGAGTACCCCTAACCT 618
```

```
Db 518 uLysGluProAlaProThrThrProLysLysProAlaPro 531
RESULT 2
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-84

Alignment Scores:
Pred. No.: 9,35e-05 Length: 1022
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
DB: 4 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-84 (1-1022)
QY 18 CCGAGCCGGCTCTCTCTGCTGCTGCTGCGCGCTGCTGCGGCGGCTAGG 77
Db 380 ProAlaProThrThrProLysGluProAlaProThrThrLysGluProSerProThr 399
QY 78 GCTCCCGCTGAGAGCC-----CGG 98
```



519 GAATGCCAGCGCTCCACGAGCACCAGGAGGTACTCCCTTGCTGCTGGAGCTGCAGAAGCT 578  
 |||||  
 563 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrLe 583  
 |||||  
 579 GCCAGATTGTCACCAACCTTGTAGTACCCCTAACCTT 618  
 |||||  
 583 uLysGluProAlaProThrThrProLysLysProAlaPro 596  
 |||||

**RESULT 4**  
 US-07-757-022B-58  
 ; Sequence 58, Application US/07757022B  
 ; Patent No. 6433142  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesner, Thomas G.  
 ; APPLICANT: Clark, Stephen C.  
 ; APPLICANT: Turner, Katherine  
 ; APPLICANT: Hewick, Rodney M.  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.  
 ; STREET: 87 CambridgePark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/757,022B  
 ; FILING DATE: 19910910  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; FILING DATE: 19-JAN-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/546,114  
 ; FILING DATE: 29-JUN-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/457,196  
 ; FILING DATE: 29-DEC-1989  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/390,901  
 ; FILING DATE: 08-AUG-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cserr, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: GI 5190  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617)876-1170  
 ; TELEFAX: (617)876-5851  
 ; INFORMATION FOR SEQ ID NO: 58:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1049 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-757-022B-58

Alignment Scores:		
Pred. No.:	9.4e-05	Length: 1049
Score:	148.50	Matches: 64
Percent Similarity:	32.48%	Conservative: 12
Best Local Similarity:	27.35%	Mismatches: 124
Query Match:	12.77%	Indels: 34
DB:	4	Gaps: 5

US-10-030-225-1\_COPY\_58\_696 (1-639) x US-07-757-022B-58 (1-1049)  
 18\_CCGAGCGCGGCTCTCTCTGCTGCTGCTGCTGGCGGCGCTGCTGCGAGCGCGCTAGG 77



```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

Alignment Scores:
Pred. No.: 9.57e-05 Length: 1140
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
Gaps: 5
DB: 4

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-104 (1-1140)

Qy 18 CCGAGCGGGCTCCTCTCGTCTGCTGCGCGCGCTGCTGGAGCGCGCTAGG 77
Db 18 CCGAGCGGGCTCCTCTCGTCTGCTGCGCGCGCTGCTGGAGCGCGCTAGG 77
Qy 78 GCTCCCGTGAAGGCC-----GCG 98
Db 78 GCTCCCGTGAAGGCC-----GCG 98
Qy 518 ThrProLysGluProAlaProThrThrLysSerAlaProThrThrLysGluPro 537
Db 518 ThrProLysGluProAlaProThrThrLysSerAlaProThrThrLysGluPro 537
Qy 99 GCTCGGGGACACCGGCTGGAGCTCAGAGCTCCAGAGTCTCA-----GCTC 152
Db 99 GCTCGGGGACACCGGCTGGAGCTCAGAGCTCCAGAGTCTCA-----GCTC 152
Qy 538 AlaProThrThrLysSerAlaProThrThrProLysGluProSerProThrThr 557
Db 538 AlaProThrThrLysSerAlaProThrThrProLysGluProSerProThrThr 557
Qy 153 CCGAGTCTAGGCTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 212
Db 153 CCGAGTCTAGGCTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 212
Qy 557 rLysGluProAlaProThrThrProLysGluProAlaProThrThrProLys 577
Db 557 rLysGluProAlaProThrThrProLysGluProAlaProThrThrProLys 577
Qy 213 GCAGGAGAGCACCACCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 270
Db 213 GCAGGAGAGCACCACCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 270
Qy 577 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaPro 597
Db 577 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaPro 597
Qy 271 -----CCAGGACACGACCCCTCCAGGACCC 299
Db 271 -----CCAGGACACGACCCCTCCAGGACCC 299
Qy 597 rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLys 617
Db 597 rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLys 617
Qy 300 AGAGTTACTCCGTTGGCTGGAGTGCAGAGCTCCGCGGATTGGCCAGCACAACTT 359
Db 300 AGAGTTACTCCGTTGGCTGGAGTGCAGAGCTCCGCGGATTGGCCAGCACAACTT 359
Qy 617 aProThrThrProLysLysLeuThrProThrThrProLysLysLeuAlaPro 637
Db 617 aProThrThrProLysLysLeuThrProThrThrProLysLysLeuAlaPro 637
Qy 360 GAGTACCCCTAACCTGATACCCAGGCTTACGCTCCCA-----GATCCTAG 407
Db 360 GAGTACCCCTAACCTGATACCCAGGCTTACGCTCCCA-----GATCCTAG 407
Qy 637 oGluLysProAlaProThrThrProGluLysLeuAlaProThrThrProGlu 657
Db 637 oGluLysProAlaProThrThrProGluLysLeuAlaProThrThrProGlu 657
Qy 408 CCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467
Db 408 CCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467
Qy 657 rProThrThrProGluGluProAlaProThrThrProLysAlaAlaProAgnThr 677
Db 657 rProThrThrProGluGluProAlaProThrThrProLysAlaAlaProAgnThr 677

Qy 468 ACACCAACATGATGTTGGACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 518
Db 468 ACACCAACATGATGTTGGACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 518
Qy 519 GAATGCCAGCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 578
Db 519 GAATGCCAGCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 578
Qy 697 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaPro 717
Db 697 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaPro 717
Qy 579 GCCAGAAATGCTCCACGCAACCTTGAGTACCCCTAACCTT 618
Db 579 GCCAGAAATGCTCCACGCAACCTTGAGTACCCCTAACCTT 618
Qy 717 uLysGluProAlaProThrThrProLysLysProAlaPro 730
Db 717 uLysGluProAlaProThrThrProLysLysProAlaPro 730

RESULT 6
US-07-757-022B-44
; Sequence 44, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-44

Alignment Scores:
Pred. No.: 9.8e-05 Length: 1270
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
```







18	CGAGCGCGCTCTCTCTGCGTGTGCTGCTGCGCGCGTGTGAGGCGCGCGCTAGG	77
	:::	
414	ProAlaProThrThrProLysGluProAlaProThrThrLysGluProSerProThr	433
	:::	
78	GCTCCCGTGAAGAAGCC-----GCG	98
	:::	
434	ThrProLysGluProAlaProThrThrLysSerAlaProThrThrLysGluPro	453
	:::	
99	GCTCCGCGGACCGCGCTGGAGCTCACGAGGCTCGCAGAGTCTCA-----GCCTC	152
	:::	
454	AlaProThrThrThrLysSerAlaPro--ThrThrProLysGluProSerProThrThr	473
	:::	
153	CCGAGATCTTAGGCTCTGAAGGAAGGAGGAGCACCTGCTCCCGAGAGCCACCT	212
	:::	
473	rLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl	493
	:::	
213	GCAGGCGAGCCACCAACATGATGCTGGACTGTCTACTGAGCCAGCAGCCATGACC--	270
	:::	
493	aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThr	513
	:::	









Db 676 uLysGluProAlaProThrThrProLysLysProAlaPro 689

Search completed: February 19, 2005, 00:24:15  
Job time : 45.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 19, 2005, 00:21:47 ; Search time 118.5 Seconds  
(without alignments)  
3529.232 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163

Sequence: 1 atcgctgcctcgccgacgg.....ataaccaggtgaccatcaag 639

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 2760536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USTO\_spool\_P/US10030225/runat\_18022005\_095903\_16366/app\_query.fasta\_1.775  
-DB=Published Applications\_AA -Qfmt=fastan -SUFFIX=rap -MINMATCH=0.1  
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10030225 @CGN 1 1 199 @runat 18022005\_095903\_16366  
-NCPU=6 -ICPU=1 NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	1053.5	90.6	658	11	US-09-939-853A-94	Sequence 94, Appl
2	879	75.6	538	15	US-10-074-978A-403	Sequence 403, App
3	430	37.0	81	9	US-09-864-761-33474	Sequence 33474, A
4	154.5	13.3	410	16	US-10-437-963-196665	Sequence 196665,
5	148.5	12.8	941	13	US-10-124-557-14	Sequence 14, Appl
6	148.5	12.8	1022	13	US-10-124-557-84	Sequence 84, Appl
7	148.5	12.8	1038	13	US-10-124-557-74	Sequence 74, Appl
8	148.5	12.8	1049	13	US-10-124-557-58	Sequence 58, Appl
9	148.5	12.8	1140	13	US-10-124-557-104	Sequence 104, Appl
10	148.5	12.8	1270	13	US-10-124-557-44	Sequence 44, Appl
11	148.5	12.8	1311	13	US-10-124-557-42	Sequence 42, Appl
12	148.5	12.8	1313	13	US-10-124-557-142	Sequence 142, App
13	148.5	12.8	1314	13	US-10-124-557-50	Sequence 50, Appl
14	148.5	12.8	1320	13	US-10-124-557-46	Sequence 46, Appl
15	148.5	12.8	1320	13	US-10-124-557-60	Sequence 60, Appl
16	148.5	12.8	1354	13	US-10-124-557-48	Sequence 48, Appl
17	148.5	12.8	1361	13	US-10-124-557-40	Sequence 40, Appl
18	148.5	12.8	1363	13	US-10-124-557-52	Sequence 52, Appl
19	148.5	12.8	1404	9	US-09-802-207-30	Sequence 30, Appl
20	148.5	12.8	1404	11	US-09-897-188-1	Sequence 1, Appli
21	148.5	12.8	1404	13	US-10-124-557-2	Sequence 2, Appli
22	148.5	12.8	1404	13	US-10-124-557-62	Sequence 62, Appl
23	148.5	12.8	1404	17	US-10-868-577A-55	Sequence 55, Appl
24	145	12.5	404	16	US-10-437-963-116565	Sequence 116565,
25	144.5	12.4	192	16	US-10-437-963-171117	Sequence 171117,
26	144.5	12.4	625	16	US-10-437-963-165015	Sequence 165015,
27	144.5	12.4	19723	15	US-10-084-846A-5	Sequence 5, Appli
28	139.5	12.0	271	16	US-10-437-963-181500	Sequence 181500,
29	136.5	11.7	481	16	US-10-437-963-176449	Sequence 176449,
30	134	11.4	420	16	US-10-437-963-201378	Sequence 201378,
31	134	11.5	655	16	US-10-437-963-141788	Sequence 141788,
32	133.5	11.5	19662	15	US-10-084-846A-6	Sequence 6, Appli
33	132.5	11.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
34	131	11.3	537	16	US-10-437-963-189969	Sequence 189969,
35	130	11.1	469	15	US-10-425-114-43328	Sequence 43328, A
36	129.5	11.1	366	16	US-10-437-963-197924	Sequence 197924,
37	129	11.1	205	16	US-10-437-963-190742	Sequence 190742,
38	129	11.1	550	11	US-10-437-963-145800	Sequence 145800,
39	129	11.1	1518	11	US-09-989-890-184	Sequence 184, App
40	128.5	11.0	423	16	US-10-437-963-158031	Sequence 158031,
41	128.5	11.0	466	16	US-10-437-963-195119	Sequence 195119,
42	128.5	11.0	582	14	US-10-156-761-11938	Sequence 11938, A
43	127.5	11.0	223	15	US-10-369-493-19842	Sequence 19842, A
44	127.5	11.0	1953	17	US-10-488-056-42	Sequence 42, Appl
45	127	10.9	19652	15	US-10-084-846A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-939-853A-94  
; Sequence 94, Application US/09939853A  
; Publication No. US20040039163A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess et al.  
; TITLE OF INVENTION: No. US20040039163A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-099  
; CURRENT APPLICATION NUMBER: US/09/939, 853A  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR FILING DATE: 2001-08-27  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: 60/267,300  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 60/269,961  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/277,337  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 159  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 94  
; LENGTH: 658

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-939-853A-94

Alignment Scores:      5.12e-70      Length:      658
Pred. No.:            1053.50      Matches:      211
Score:                70.67%      Conservative: 1
Percent Similarity:   70.33%      Mismatches:  1
Best Local Similarity: 90.58%      Indels:      87
Query Match:         11          Gaps:        1
DB:

US-10-030-225-1_COPY_58_696 (1-639) x US-09-939-853A-94 (1-658)

QY 1 ATGCGTGGCTCCGCGACCGAGCGGCTCTCTCGTGGTGGTGGCGGCGCTG 60
D 1 MetArgAlaLeuArgAspArgAlaGlyLeuLeuLeuCysValLeuLeuAlaLeu 20
QY 61 CTGAGGCGGCGCTAGGCTCCCGTGAAGAGCGGGCTCCGCGGACCGCGCTGGG 120
D 21 LeuGluAlaLeuGlyLeuProValLysProArgLeuArgGlyProArgProGly 40
QY 121 AGCTTCAGGAGCTCGCAGAG----- 141
D 41 SerLeuThrArgLeuAlaGluValSerGlyGlyThrGlyLeuArgSerAlaLeuSer 60
QY 141 ----- 141
D 61 ValProProGlnProAlaGlySerSerArgAlaGlySerGlyThrGlyThrHisThr 80
QY 141 ----- 141
D 81 GlySerAspProProMetGluArgGlyAlaGlyAlaGlyLysLeuProAspThrGly 100
QY 141 ----- 141
D 101 ArgCysProValThrGluGlySerThrValGlnLeuLeuAlaProTrpAsnAlaAlaAsp 120
QY 142 -----GCTCAGCTCCCGCAT 159
D 121 ValHisSerHisGlyAspLysAspSerGlnThrCysIleArgValSerAlaSerProasp 140
QY 160 CTTAGGCTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
D 141 ProArgProLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 160
QY 220 GAGCACACCAACATGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 279
D 161 GluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThrProGlyAsn 180
QY 280 ACCACCTCCAGAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 339
D 181 AlaThrProArgThrProGluValThrProLeuArgLeuGluGlnLysLeuPro 200
QY 340 GGATTGGCCAGCACACCTTGAGTACCCCTTAACCTGTATACCCAGGCTTCAGCCTCCCA 399
D 201 GlyLeuAlaAsnThrThrLeuSerThrProAsnProAspThrGlnAlaSerAlaSerPro 220
QY 400 GATCCTAGGCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
D 221 AspProArgProLeuArgGluGluGluAlaArgLeuLeuProArgThrHisLeuGln 240
QY 460 GCAGACTACCAACATGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 519
D 241 AlaGluLeuHisGlnHisGlyCysTrpThrValThrGluProAlaAlaLeuThrProGly 260
QY 520 AATGCCAGCGCTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 579
D 261 AsnAlaThrProProArgThrGlnGluValThrProLeuLeuGluGlnLysLeu 280
QY 580 CCAGATTGGTCCAGCAACCTTGATGATCCCTTAACCTGATACCCAGGTGACCATCAAG 639
D 281 ProGluLeuValHisAlaThrLeuSerThrProAsnProAspAsnGlnValThrIleLys 300
```

```
RESULT 2
US-10-074-978A-403
; Sequence 403, Application US/10074978A
; Publication No. US20040010119A1
; GENERAL INFORMATION:
; APPLICANT: Leite, Mario
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Fernandes, Elma
; APPLICANT: Li, Li
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Liu, Xiahong
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Patturajan, Meera
; APPLICANT: Bialock, Angela
; APPLICANT: Ballinger, Robert
; APPLICANT: Vernet, Corine
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Gusev, Vladimir
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter S
; APPLICANT: Ellerman, Karen
; APPLICANT: Heyes, Melvin P
; APPLICANT: Herrman, John
; APPLICANT: Pena, Carol E A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Moore, NO. US20040010119A1lle
; APPLICANT: Shenoy, Suresh
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, Dave
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-269
; CURRENT APPLICATION NUMBER: US/10/074,978A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: 60/268,221
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/335,109
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/312,284
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/268,496
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/276,703
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/330,293
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/322,127
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/280,899
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/310,797
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/268,646
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 403
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-978A-403

Alignment Scores:      4.66e-57      Length:      538
Pred. No.:

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33474
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF111168.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.75
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EST HUMAN HIT: BE281519.1, EVALUAE 1.30e+00
; OTHER INFORMATION: EST HUMAN HIT: BE731965.1, EVALUAE 2.00e-31
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUAE 5.30e+00
US-09-864-761-33474

Alignment Scores:
Pred. No.: 9.34e-24 Length: 81
Score: 430.00 Matches: 79
Percent Similarity: 98.77% Conservative: 1
Best Local Similarity: 97.53% Mismatches: 1
Query Match: 36.97% Indels: 0
DB: 9 Gaps: 0

US-10-030-225-1_COPY_58_696 (1-639) x US-09-864-761-33474 (1-81)
QY 142 GTCTCAGCCTCCCGAGATCCTAGGCCTCTGAAGGAAGAGGAGGACCACTGCTCCCC 201
DB 1 ValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeuPro 20
QY 202 AGAACCCACTGCAGGAGGAGCCACACCAACATGGATGCTGGACTGTCACTGAGCAGCA 261
DB 21 ArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAla 40
QY 262 GCATGACCCCGAGCAACACACCCCTCCAGGAGCCCGAGGTTACTCCGTTGGCGGTG 321
DB 41 AlaMetThrProGlyAsnAlaThrProProArgThrProGluValThrProLeuArgLeu 60
QY 322 GAGCTGCAGAACTCGCGGATTCGGCAGGAGCCACCACTTGTAGTACCCCTGATACC 381
DB 61 GluLeuGlnLysLeuProGlyLeuAlaAsnThrThrLeuSerThrProAsnProAspThr 80
QY 382 CAG 384
DB 81 Gln 81

```

```

Alignment Scores:
Pred. No.:      0.00312      Length:      410
Score:          154.50      Matches:      75
Percent Similarity: 41.26%    Conservative:  17
Best Local Similarity: 33.63%  Mismatches:   57
Indels:         13.28%      Gaps:         75
Query Match:    16
DB:

US-10-030-225-1_COPY_58_696 (1-639) x US-10-437-963-196665 (1-410)

QY      9  GCTCCGGCAGCCGAGCGGGCTCTCTCTGCTGCTGCTGTGG----- 52
DB     90  AlaProLeuPtoThraLaSerProProArgAlaAlaArgAlnProThraLaVal 109
QY     53  CGGCGCTGCTGGAGCGCGCGCTAGGCTCCCTGTGAAGAGCGCGCTCCGGCAGCCAC 112
DB    110  HisArgArgTtpArgArgArg--SerGlnProValArgProArg-AlaProLeuThraL 128
QY    113  GGCCTGGGAGGCTCAGCAGGCTGCAGAGTCTCAGCCTCCCGAGATCTCTAGCGCTCTGA 172
DB    128  aSerProProProSerSerAlaAlaArg-----AlaAlaL 140
QY    173  AGGAAGAGGAGGAGGAGCCACTGCTCCCGAGAACCCACTGCAGGAGCAGA---GCCACACC 229
DB    140  aHisArgGlnSerGly--ProAlaAlaArgSerProProAlaHisArgGlnSerAlaPr 159
QY    230  AACATGGATGTGGACTGTCTACTGAGCCAGCAGCATGACCCAGGACACACACCCCTTC 289
DB    159  0thr-----ArgArgSerArgSerProSerProHisProProIn 172
QY    290  CCAGGACCCAGAGTTACTCCGTTGGGCTGGAGCTGCAGAAAGCTGCCGGGATTGGCCCA 349
DB    172  thisProProArgGly-AlaPro-CysSerProGlyGluAspSerCys---ThrTtpPro 190
QY    350  GCACACACCTTGATACCCCTTAACCTGTATACCCAGGCTTCAGCTCCCGAGATCTTAGGC 409
DB    191  Ser-----ProCysThrProLeuProArgHisGln----- 200
QY    410  CTCTGAGGAGAGGAGGAGGAGCAGACTGCTCCCGAGAACCCACCTCGAGGCAGAGCTAC 469
DB    201  -----ArgArgHisGlyCysAlaProAlaProSerAlaAlaAlaAlaIn 216

```

Alignment Scores:		
Pred. No.:	0.00328	941
Score:	148.50	Matches: 64
Percent Similarity:	34.4%	Conservative: 12
Best Local Similarity:	27.3%	Mismatches: 124
Query Match:	12.7%	Indels: 34
DB:	13	Gaps: 5

١٠







**Alignment Scores:**







## RESULT 14







Result No.	Score	Query Match	Length	DB ID	Description
1	150.5	12.9	460	2	T33110
2	136.5	11.7	880	2	D89756
3	134	11.5	553	2	E84682
C	133.5	11.5	383	2	S32975
	133.5	11.4	2639	2	T33128
	128	11.0	395	2	H75457
	127.5	11.0	294	2	A37232
	126.5	10.9	825	1	ED8BXD
9	126	10.8	377	2	A48018
C	126	10.8	1428	2	T08852
	126	10.8	2849	2	T51023
11	125	10.7	3149	1	Q08E8
C	123.5	10.5	446	2	T45525
	123.5	10.6	1323	2	I79557
14	123.5	10.6	1323	2	I79557
					hypothetical prote
					protein T23E7_2b [
					hypothetical prote
					gene BCRF2 Chinese
					fibroin - fibroin
					hypothetical prote
					mucin, tracheal (A
					immediate-early pr
					mucin 7 precursor,
					lustrin A - Califo
					hypothetical prote
					BP1F1 protein - hu
					WSC4 homolog [impo
					N-methyl-D-asparta











1882 -----AlaGluGluPProGlnProAlaIuYbSerGluAlaAaIaProGProAlaAbnGlnPr 1899

QY 196 -----CTCCCCAGAACCCCATGCGAGGCAGGCACACACCAATGGATGCTGGAC 245

Db 199 oPheSerArgPheGlnValThrProIleGlnAlaSerProValLeuHis----- 1715

QY 246 TGTCACCTGAGCCAGCAGCATGACCCAGGCAACACACCCCTCCAGGACCCAGAGGT 305

Db 1716 -----ThrLeuValGlnArgSerProThrVa 1724

QY 306 TACTCCGTTGCGGCTGGAGCTGCAGAAAGTGC CGGATTTGGCCAGCACA----- 354

Db 1724 lMetProAlaProLeuProProAlaProValAlaAaIaSerAlaSerAlaAaIaProSerPr 1744

QY 355 ----ACCTTGAGTACCCCTAAACCTGATACCCAGGCTTCAGCCTCCCCAGATCCTAGGCC 410

Db 1744 oGlnThrProSerArgProGlyProAlaValSerGlnThrMetSerProValProHisPr 1764

QY 411 TCTGAGGGAA-----GAGGAGGAGGCAGCTGCTCCCCAG 446

Db 1764 oLeuArgGlnProThrAlaThrPheValPheThrGluArgGluGlyGluProIleProVa 1784

QY 447 AACCACCTGCAGGCAGAGCATACCAACATGGATGTTGGAGCTGTCACTGAGCCAGCAGC 506

Db 1784 lSerGlnProProAlaGlnThr---GlnHisGlnProValArgIleSerGlnIuysThrAl 1803

QY 507 CCTGAGCCAGGAAATGCCAGCCT----- 531

Db 1803 aProValProSerSerSerMetProSerAlaSerGluAlaMetProArgSerAlaGlyTr 1823

QY 532 ----CCAGGACCCAGGAGGTACTCCCTTGCTGTGCGAGCTGCAGAAAGTGCACGAATT 587

Db 1823 pIleProGluAlaAaenArgProThrProLeuLeuSerGlnGlnHisGluLeuArgGluVa 1843

QY 588 G 588

Db 1843 i 1843

RESULT 12

QBB8

BPLF1 protein - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus

C:Date: 25-Feb-1985 #sequence revision 25-Feb-1985 #text\_change 09-Jul-2004

C:Accession: G93065; A03747; S32993

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus genome.

A:Reference number: A93065; MUID:850357113; PMID:6092825

A:Accession: G93065

A:Molecule type: DNA

A:Residues: 1-3149 <BAN>

A:Cross-references: UNIPROT:P03186; EMBL:V01555; NID:G59074; PIDN:CAA24839.1; P:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson

Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667; PMID:6087149

A:Contents: annotation; protein coding region

C:Superfamily: human herpesvirus 4 BPLF1 protein

```

Alignment Scores:
Pred. No.:          0.12          Length:          3149
Score:             125.00         Matches:         65
Percent Similarity: 39.57%       Conservative:    28
Best Local Similarity: 27.66%    Mismatches:     85
Query Match:       10.75%       Indels:         57
DB:                1           Gaps:           15

US-10-030-225-1_COPY58_696 (1-639) x Q0BE8 (1-3149)

;
Qy      12  CCGCGACCGAGCGGGCTCCTCTCTGCTGCTGCTGCGGGCGGCGCTCTGAGGCGGC 71
      |||||
Db      395  PrcAlaProSerThrProPro-----ArgAlaSerSerGtValaAla----- 408
      |||||

```









```

Qy      515 CAGGGAATGCCACGCTCCACGAGCCACGAGAGTTACTCCCTTCTGCTGGAGCTGCAGA 574
Db      1101 oSerAspSerGluAspSerGluSerLeuGlyGlyAlaSerLeu-----GlyGlyLeuGl 1119
Qy      575 A-----GCTGCCAGAAATGGTCCA 593
Db      1119 uProTTrpPheAlaAspPheProTyrProTyrAlaGluArgLeuGlyPro 1136

```

Search completed: February 19, 2005, 00:23:05  
Job time : 51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 20:52:54 ; Search time 426 Seconds  
(without alignments)  
8879.612 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 639  
Sequence: 1 atgcgtgcgtccgcagccg.....ataaccaggtagaccatcaag 639

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	2981	4	Aaf29348 Human pro
2	639	100.0	2981	5	Aaf93820 Human cdn
3	542.6	84.9	750	5	Aaf94019 Primer sp
C 4	261.8	41.0	497	4	Aai10232 Probe #16
C 5	261.8	41.0	497	4	Aba51868 Human foe
C 6	261.8	41.0	497	4	Aai31480 Probe #16
C 7	261.8	41.0	497	4	Aba21690 Probe #15
C 8	261.8	41.0	497	4	Aak25606 Human bon
C 9	261.8	41.0	497	4	Aak00167 Human bra
C 10	261.8	41.0	497	4	Abs25182 Human liv
C 11	261.8	41.0	497	5	Aai00170 Probe #16
C 12	261.8	41.0	497	6	Abs00175 Human gen
C 13	239.8	37.5	243	4	Aai19514 Probe #94
C 14	239.8	37.5	243	4	Aba64531 Human foe
C 15	239.8	37.5	243	4	Aai44705 Probe #13
C 16	239.8	37.5	243	4	Aba46659 Human bre
C 17	239.8	37.5	243	4	Aba31664 Probe #10
C 18	239.8	37.5	243	4	Aak38709 Human bon
C 19	239.8	37.5	243	4	Aak12982 Human bra
C 20	239.8	37.5	243	4	Abs38283 Human liv

C 21	239.8	37.5	243	5	AAI05236	Aai05236 Probe #52
C 22	239.8	37.5	243	6	ABS12780	Abs12780 Human gen
C 23	141	22.1	522	10	ADFS8409	Adfs8409 Human pol
C 24	91.8	14.4	1128	6	ABQ30568	Abq30568 Oligonuc1
C 25	91.8	14.4	1128	6	ABQ30569	Abq30569 Oligonuc1
C 26	85.6	13.4	1128	6	ABQ30566	Abq30566 Oligonuc1
C 27	85.6	13.4	1128	6	ABQ30567	Abq30567 Oligonuc1
C 28	44.8	7.0	1047	11	ABD10430	Abd10430 Pseudomon
C 29	44.8	7.0	1464	11	ABD10367	Abd10367 Pseudomon
C 30	44.8	7.0	1494	11	ABD10513	Abd10513 Pseudomon
C 31	44.8	7.0	2704	12	ADQ64839	Adq64839 Novel hum
C 32	42.6	6.7	3018	4	AAH78274	Aah78274 Coding se
C 33	42.6	6.7	3391	4	AAH78273	Aah78273 Nucleotid
C 34	42.6	6.7	5968	4	AAS27841	Aas27841 DNA encod
C 35	42.6	6.7	5968	4	AAS35091	Aas35091 DNA #41 e
C 36	42.6	6.7	5968	4	AAK80295	Aak80295 Human imm
C 37	42.6	6.7	5968	10	ADB94644	Adb94644 Novel hum
C 38	42.6	6.7	5968	10	ADC46533	Adc46533 Human neo
C 39	42.2	6.6	815	3	AAA96721	Aaa96721 Polynucle
C 40	42.2	6.6	3334	13	ADR08320	Adr08320 Full leng
C 41	41.8	6.5	15872	2	AAT68715	Aat68715 Streptomy
C 42	41.6	6.5	4705	8	ACA61566	Aca61566 Streptomy
C 43	41.6	6.5	4736	8	ACA61565	Aca61565 Streptomy
C 44	41.6	6.5	8174	6	AAI44951	Aai44951 Equine he
C 45	41.6	6.5	149158	12	ADP74211	Adp74211 Equine he

## ALIGNMENTS

## RESULT 1

AAF29348

ID AAF29348 standard; cdna; 2981 BP.

XX AAF29348;

AC AAF29348;

DT 20-APR-2001 (first entry)

XX Human proliferation differentiation factor cdna sequence.

DE Human proliferation differentiation factor; haematopoietic function; ss.

XX Homo sapiens.

XX WO200104312-A1.

XX 18-JAN-2001.

XX 06-JUL-2000; 2000WO-JP004514.

XX 08-JUL-1999; 99JP-00194179.

XX 18-OCT-1999; 99US-0159586P.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

XX WPI: 2001-138354/14.

XX P-PSDB: AAB49765.

XX Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cdna library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.

XX Claim 1; Page 36-41; 49pp; Japanese.

XX This invention relates to polynucleotide sequence PSEC137, which encodes a human proliferation differentiation factor protein. Included in the invention is a vector containing the cdna sequence, a transformant containing the vector, and a process for producing the protein. The protein together with its encoded DNA and antisense DNA are used in drugs to maintain the haematopoietic function. The present sequence represents cdna encoding the proliferation differentiation factor protein























**This Page Blank (uspto)**

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	44.8	7.0	1047	4	US-09-252-991A-9034	Sequence 9034, Appl
C 2	44.8	7.0	1464	4	US-09-252-991A-8971	Sequence 8971, Appl
C 3	44.8	7.0	1494	4	US-09-252-991A-9117	Sequence 9117, Appl
C 4	43.6	6.8	1494	4	US-09-502-540-6461	Sequence 6461, Appl
C 5	43.6	6.8	1664	4	US-09-502-540-452	Sequence 452, Appl
C 6	42.2	6.6	815	3	US-09-383-586-9	Sequence 9, Appl
C 7	42.2	6.6	815	4	US-09-623-038A-9	Sequence 9, Appl
C 8	41.2	6.4	732	4	US-09-252-991A-8860	Sequence 8860, Appl
C 9	41	6.4	2283	3	US-09-252-991A-3298	Sequence 3298, Appl
C 10	39.8	6.2	15872	3	US-09-105-537-1	Sequence 1, Appl
C 11	39.8	6.2	15872	4	US-09-091-609-1	Sequence 1, Appl
C 12	39.8	6.2	15872	4	US-09-091-609-3	Sequence 3, Appl
C 13	39.8	6.2	43280	2	US-08-804-2272-1	Sequence 1, Appl
C 14	39.4	6.2	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 15	39.8	6.1	40586	4	US-09-949-016-16965	Sequence 16965, A
C 16	38.6	6.0	1638	4	US-10-237-551-119	Sequence 119, Appl
C 17	38.6	6.0	1638	4	US-10-237-551-148	Sequence 148, Appl
C 18	38.6	6.0	1644	4	US-10-237-551-213	Sequence 213, Appl
C 19	38.6	6.0	1644	4	US-10-237-551-214	Sequence 214, Appl
C 20	38.6	6.0	154746	4	US-09-827-688-8	Sequence 8, Appl
C 21	38.4	6.0	621	4	US-09-252-991A-6380	Sequence 6380, Appl
C 22	38.4	6.0	933	4	US-09-252-991A-6169	Sequence 6169, Appl
C 23	38.4	6.0	1350	4	US-09-252-991A-6012	Sequence 6012, Appl
C 24	38.2	6.0	1393	2	US-08-765-875-1	Sequence 1, Appl
C 25	38.2	6.0	1393	3	US-08-795-671-1	Sequence 1, Appl
C 26	38.2	6.0	1393	4	US-09-454-540-1	Sequence 1, Appl
C 27	38.2	6.0	1393	4	US-09-626-896-24	Sequence 24, Appl

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

, FILE REFERENCE: 107196.136
, CURRENT APPLICATION NUMBER: US/09/252,991A
, CURRENT FILING DATE: 1999-02-18
, PRIOR APPLICATION NUMBER: US 60/074,788
, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,190
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 8971
, LENGTH: 1464
, TYPE: DNA
, ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8971

```

	Query Match	7.0%;	Score 44.8;	DB 4;	Length 1464;
	Best Local Similarity	50.0%;	Pred. No. 0.11;		
	Matches 112;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;
Qy	7	GCGCTCCGACCGAGCGGGCTCTCTCTGTCGTCTGCTGTGCGGGCGCTCTCTGGAG	66		
Db	685	GAGCGCGCAAGCTGGCGACGGATTTCCACCGCGTCTGTTGCCCGCTCTCTGGC	744		
Qy	67	GCGCGCGTAGGGTCTCCCGTGAGAAAGCGCGGGCTCCCGCGAACACCGCTCTGGAGCGTCTC	126		
Db	745	CAGCGCAGGCGCGCGCGAGGAACAGCGCCCTTCTCCGGCGCCCGCGCAAGCGACTG	804		
Qy	127	ACGAGGCTCCAGAGGTCTCAGCTCTCCAGATCCTAGGCTCTGAGGAGAGAGAGGAG	186		
Db	805	CACGCTCTGCCAGGGCGACTCCGCGAGGATCAGTGTGGCTCTCGCGCGACGTAGGGCAG	864		
Qy	187	GCACCACTGTCTCCAGAACCCACTCTCAGGCAGAGCCACACCA	230		
Db	865	GCTGACGTAGCGCGCGCTCTCTTCGCGCAAGCGCAGCAACGGCA	908		

## RESULT 3

RESULT 3  
US-09-252-991A-9117/c  
; Sequence 9117, Application US/09252991A

```

1  Patent No. 8531793
2  GENERAL INFORMATION:
3  APPLICANT: Marc J. Rubenfield et al.
4  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
5  TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
6  FILE REFERENCE: 107196.136
7  CURRENT APPLICATION NUMBER: US/09/252,991A
8  CURRENT FILING DATE: 1999-02-18
9  PRIOR APPLICATION NUMBER: US 60/074,788
10 PRIOR FILING DATE: 1998-02-18
11 PRIOR APPLICATION NUMBER: US 60/094,190
12 PRIOR FILING DATE: 1998-07-27
13 NUMBER OF SEQ ID NOS: 33142
14 SEQ ID NO 9117
15 LENGTH: 1494
16 TYPE: DNA
17 ORGANISM: Pseudomonas aeruginosa
18 US-09-252-991A-9117

```

Query Match	7.0%;	Score 44.8;	DB 4;	Length 1494;
Best Local Similarity	50.0%;	Pred. No. 0.11;		
Matches 112; Conservative	0;	Mismatches 112;	Indels 0;	Gaps 0;
QY	7	GC	CTCGGCACGAGCGGGCTCTCTCTGCGTGTCTGTGCGCGCTGCTGGAG	66
Db	879	GAGCGCGCAAGCTGGCGACGGA	TTTCCACCGCGTCTCTGTGTTGCCGCTCTCGC	820
QY	67	GC	CGCGCTAGGGGTCCCGCTGAAGACCGCGCTCCGCGGACACACGGCTTGGAGGCTC	126
Db	819	CAGCGCAGCGCCCGCGAGGACAGCGCTTCTCCGGGCGCGCGGCAAGCACTG	760	
QY	127	AC	GAGGCTCGCAGGTCTCAGCCTCTCCAGATCTTAGGCTCTGAAGGAGGAGGAG	186
Db	759	CACGCGTCTCCCAGGGCGGACCTCCGCGAGGATCAGTGGCGTCCGGCGGACGTAGGCGAG	700	

Qy 187 GCACCACTGCTCCCTCCAGAACCCACCTGCAGGCAGGCCACCA 230  
Dδ 699 GCTGACGTAGGCGCGCCCTCTCTCGCCACGGCGAGCAACGGCA 656

## RESULT 4

```

US-09-902-540-6461
; Sequence 6461, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6461
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/key: unsure
; LOCATION: (1)..(1494)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-6461

```

Query Match  
6.8%; Score 43.6; DB 4; Length 1494;

	QY	Db	QY	Db
Best Local Similarity	46.3%			
Pred. No.	0.22			
Mismatches	244			
Indels	5			
Gaps	27			
Matches	215			
Conservative	0			
	16	550	76	610
	GACGAGCCGGGCTCTCTCTCGTGTCTCTCTCTGCGCGGCTGCTCTGAGCGCGGCTA	GGAGCGAGCCTCATCCACTCATGCTCCAGGTGCGGGTGTGGAGACCGGCTGGT	GGCTCCCCGTGAAGACCGCGGCTCCGCGGACACGGCTGGAGCCCTCACGAGGCTC	CTCTCTGCGATGCGAACCTTGGGCTCATTTGGGAGGAGGCCGCGAGCCTCCGCGCCCTC
	75	609	135	669

QY	136	GCAGAGTCTCAGCCTCCCAGATCTTAGCCCTCTGAAGCAAGAGGAGGAGCCACCACTG	193
Db	670	AACCAAGCATCGCTTAC---GTGCCCAAGTTGACCTGTACCTGGACGCCACCGCGAG	726
QY	196	CTCCCAAGAACCACTGCGAGGAGAGCCACACATCATGATGCTGGACTGTCACTAG	255
Db	727	TTCCACGGGCGCAACGAGTGTCCCGAGCGCGAGCGGGTGGCCAACTGCTGGTGGTGGAG	786
QY	256	CGAGCAGCATACCCAGGCAACACACACCCCTCCAGGACCCACAGAGGTTACTCCGTTG	315
Db	787	CCCAACGGCAGAGCACTTCTTCCACCGCGGAGCGGAGGCGCGGACACGCCACG	846
QY	316	CGGCTGGAGTGCAGAAGTTCGCGGATTCGGCCAGCAACAACCTTGAGTACG--CCTAAC	373
Db	847	CGCATCTCGTGCAGCTGAGCTCCGCGCGAGCGGCGCGCTGAAATCACCGGCTCCAGC	906
QY	374	CTGATACCCAGGCTTCAGCCTCCCCAGATCCTTAGGCCTCTGAGGAAGAGGAGGAGCAC	433
Db	907	TCCGTGGGGGAGCAGCGGCGCCGAGTACCGCGCGCCTTACCGCCGAGGCCACGCGC	966
QY	434	GACTGTCCCCAGAACCCACTTCGAGGCAGAGCTACACCAACAT	477
Db	967	AAGTCCACCTTCAGGCGCGCTTCGGCGCAGAGCTTCCCGGCT	1010

## RESULT 5

US-09-902-540-452/c  
; Sequence 452, Application US/09902540  
; Patent No. 6833447







; ORGANISM: Streptomyces venezuelae		; Db	
; FEATURE:		; Qy	
; NAME/KEY: CDS		; Db	
; LOCATION: (20)...(13909)		; Qy	
US-09-091-609-1		; Db	
Query Match		; Qy	
Best Local Similarity 6.2%; Score 39.8; DB 4; Length 15872;		; Db	
Matches 134; Conservative 0; Mismatches 157; Indels 0; Gaps 0;		; Qy	
; 42 GCTGCTGTCGGCGGCTGCTGGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCT 101		; Db	
; 8773 GCTGCTGTCGAGGTGTCTGGAGGCGCTGGAACGCGGGCATGACCGTCTCTGCT 8832		; Qy	
; 102 CCGGGACACCGGCTGGAGCGCTCACGAGGCTCGCAGAGGTCTCAGCTCCCGATGCC 161		; Db	
; 8833 GCGGGACGCGGCTGTGTCTAGTGGGCGCGGCGCATCGTAGGCTCCGATCC 8892		; Qy	
; 162 TAGGCTCTGAAGAGAGGAGGACACCTGCTCCCGAGAACCCACCTGCGAGGAGA 221		; Db	
; 8893 CCGGCTGTGCCGAGGCTCGAGGGCTATCTGCTGACCGGACGCGCGCGGTGAT 8952		; Qy	
; 222 GCACACCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 281		; Db	
; 8953 GTCCGGCGCATCTCTACGCGCTCGGTCTCGAAGGACCGTCCATGAGCGTGGAGCGC 9012		; Qy	
; 282 CACCCCTCCAGGACCCAGAGGTTACTCTCGTTCGGCTGGAGCTGCAGAA 332		; Db	
; 9013 CTGCTCTCTCTGCTGGTGGCGCTGCATCTGCGGTAGCGGCGTGGCGCA 9063		; Qy	
RESULT 12		; Db	
US-09-091-609-3		; Qy	
; Sequence 3, Application US/09091609		; Db	
; Patent No. 6600029		; Qy	
; GENERAL INFORMATION:		; Db	
; APPLICANT: SHERMAN, DAVID H.		; Qy	
; APPLICANT: WILLIAMS, MARK D.		; Db	
; APPLICANT: XUE, YONGQUAN		; Qy	
; TITLE OF INVENTION: METABOLIC ENGINEERING OF		; Db	
; FILE REFERENCE: 600.297US2		; Qy	
; CURRENT APPLICATION NUMBER: US/09/091.609		; Db	
; EARLIER FILING DATE: 1998-06-19		; Qy	
; EARLIER APPLICATION NUMBER: PCT/US96/20119		; Db	
; EARLIER FILING DATE: 1996-12-18		; Qy	
; EARLIER APPLICATION NUMBER: 60/008,847		; Db	
; EARLIER FILING DATE: 1995-12-19		; Qy	
; NUMBER OF SEQ ID NOS: 4		; Db	
; SOFTWARE: FastSeq for Windows Version 3.0		; Qy	
; SEQ ID NO 3		; Db	
; LENGTH: 15872		; Qy	
; TYPE: DNA		; Db	
; ORGANISM: Streptomyces venezuelae		; Qy	
; FEATURE:		; Db	
; NAME/KEY: CDS		; Qy	
; LOCATION: (14148)...(15824)		; Db	
US-09-091-609-3		; Qy	
Query Match		; Db	
Best Local Similarity 6.2%; Score 39.8; DB 4; Length 15872;		; Qy	
Matches 134; Conservative 0; Mismatches 157; Indels 0; Gaps 0;		; Db	
; 42 GCTGCTGTCGGCGGCTGCTGGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCT 101		; Qy	
; 8773 GCTGCTGTCGAGGTGTCTGGAGGCGCTGGAACGCGGGCATGACCGTCTCTGCT 8832		; Db	
; 102 CCGGGACACCGGCTGGAGCGCTCACGAGGCTCGCAGAGGTCTCAGCTCCCGATGCC 161		; Qy	
; 8833 GCGGGACGCGGCTGTGTCTAGTGGGCGCGGCGCATCGTAGGCTCCGATCC 8892		; Db	
; 162 TAGGCTCTGAAGAGAGGAGGAGGACACCTGCTCCCGAGAACCCACCTGCGAGGAGA 221		; Qy	
; 8953 GTCCGGCGCATCTCTACGCGCTCGGTCTCGAAGGACCGTCCATGAGCGTGGAGCGC 9012		; Db	
; 282 CACCCCTCCAGGACCCAGAGGTTACTCTCGTTCGGCTGGAGCTGCAGAA 332		; Qy	
; 9013 CTGCTCTCTCTGCTGGTGGCGCTGCATCTGCGGTAGCGGCGTGGCGCA 9063		; Db	
RESULT 13		; Qy	
US-08-804-227C-1		; Db	
; Sequence 1, Application US/08804227C		; Qy	
; Patent No. 5876991		; Db	
; GENERAL INFORMATION:		; Qy	
; APPLICANT: DeHoff, Bradley S.		; Db	
; APPLICANT: Kuhstoss, Stuart A.		; Qy	
; APPLICANT: Rostock, Paul R., Jr.		; Db	
; APPLICANT: Sutton, Kimberly L.		; Qy	
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES		; Db	
; NUMBER OF SEQUENCES: 15		; Qy	
; CORRESPONDENCE ADDRESS:		; Db	
; ADDRESSEE: THOMAS G. PLANT 1501		; Qy	
; STREET: LILLY CORPORATE CENTER		; Db	
; CITY: INDIANAPOLIS		; Qy	
; STATE: IN		; Db	
; COUNTRY: USA		; Qy	
; ZIP: 46285		; Db	
; COMPUTER READABLE FORM:		; Qy	
; MEDIUM TYPE: Floppy disk		; Db	
; COMPUTER: IBM Compatible		; Qy	
; OPERATING SYSTEM: MS-DOS		; Db	
; SOFTWARE: ASCII(DOS) Text only		; Qy	
; CURRENT APPLICATION DATA: US/08/804,227C		; Db	
; FILING DATE: February 21, 1997		; Qy	
; CLASSIFICATION: 435		; Db	
; ATTORNEY/AGENT INFORMATION:		; Qy	
; NAME: Plant, Thomas, G.		; Db	
; REGISTRATION NUMBER: 35,784		; Qy	
; REFERENCE/DOCKET NUMBER: X-8231		; Db	
; TELECOMMUNICATION INFORMATION:		; Qy	
; TELEPHONE: 317-276-2459		; Db	
; INFORMATION FOR SEQ ID NO: 1:		; Qy	
; SEQUENCE CHARACTERISTICS:		; Db	
; LENGTH: 43280 base pairs		; Qy	
; TYPE: nucleic acid		; Db	
; STRANDEDNESS: single		; Qy	
; TOPOLOGY: linear		; Db	
; MOLECULE TYPE: DNA (genomic)		; Qy	
; FEATURE:		; Db	
; NAME/KEY: CDS		; Qy	
; LOCATION: 816..14234		; Db	
; FEATURE:		; Qy	
; NAME/KEY: CDS		; Db	
; LOCATION: 14351..19945		; Qy	
; FEATURE:		; Db	
; NAME/KEY: CDS		; Qy	
; LOCATION: 20010..31199		; Db	
; FEATURE:		; Qy	
; NAME/KEY: CDS		; Db	
; LOCATION: 31232..36067		; Qy	
; FEATURE:		; Db	
; NAME/KEY: CDS		; Qy	
; LOCATION: 36249..41774		; Db	
; US-08-804-227C-1		; Qy	
Query Match		; Db	
Best Local Similarity 6.2%; Score 39.8; DB 2; Length 43280;		; Qy	
Matches 134; Conservative 0; Mismatches 157; Indels 0; Gaps 0;		; Db	
; 42 GCTGCTGTCGGCGGCTGCTGGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCT 101		; Qy	

Db 36635 GCTGCTGAGGTGCTCTGGAGGCGCTGGAACGCGCGGCATCGACCGTCTCGCT 36694  
Qy 102 CCGGGACACCGCTGGAGCCTCACAGGCTCGCAGAGTCTCAGCCTCCCGATCC 161  
Db 36695 GCGGGACGCGCGGTGTGTCTACGTGGCGCGCGCACGGCTGTACGCTCGATCC 36754  
Qy 162 TAGGCTCTGAAGGAAGAGGAGGACCACTGCTCCCCAGAACCCACCTGCGAGGAGA 221  
Db 36755 CCGGCTGTGCTCCGAGGCTCGAGGCTATCTGTGACCGCGCAGCGCGGTGAT 36814  
Qy 222 GCCACACCAATGATCTGCACTGTGACGACGACGACCATGACCCCGAGCAACAC 281  
Db 36815 GTCCGGCGCATCTCTACGGGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGC 36874  
Qy 282 CACCCCTCCAGGACCCAGAGTTACTCCGTTGGGTGAGCTGCAGAA 332  
Db 36875 CTGCTCTCTCGTGTGGGCTGTCATCTGGCGGTACGGCGCTGCGGCA 36925

## RESULT 14

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 6.2%; Score 39.4; DB 3; Length 4403765;  
Best Local Similarity 52.8%; Pred. No. 19;  
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
Qy 28 CTCCTCTCTGCTGCTGCTGCTGCGCGCTGCTGAGGCGCGCTAGGCTCCCGTG 87  
Db 285911 CCCCTGCCAAGTACACAAAGTGGGCGGTGATCCGCTGCCCTGTGCTGGGCTTG 285852  
Qy 88 AAGAGCGCGGCTCCGCGACCAAGGCTGGAGCCTCACAGGCTCGCAGAGGTCTCA 147  
Db 285851 GCGAGCTGCTGAGCGGAGTTCCGCTGCCAGCAGCGCGCGGCGGTGCTGCGC 285792  
Qy 148 GCCTCCCGAGATCTAGGCTCTGAGGAAGAGGAGGAGGC 188  
Db 285791 GCGTTCGCGCACCCCGAGCGGACAAAGCAGTAGCGGTGGC 285751

## RESULT 15

US-09-949-016-16965  
; Sequence 16965, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16965  
; LENGTH: 40586  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16965

Query Match 6.1%; Score 38.8; DB 4; Length 40586;  
Best Local Similarity 51.8%; Pred. No. 8.2;  
Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
Qy 50 TGGCGGCGCTGCTGAGGCGCGCTAGGGCTCCCGTGAAGAAGCCGCGGCTCCCGGAC 109  
Db 2143 TGGCGGCGCTCCCGAGACGAGCTCAGTTCCCGAGGAGGCTCGGAGCTCGGGCAG 2202  
Qy 110 CACGCGCTGGAGGCTCACAGGCTTCGAGAGGTCTCAGCCTCCCGAGATCCTAGGCTC 169  
Db 2203 AGCCTGCGCGCGCGCCCTGGAGGACGCGCGGATGAGCCTCCCGCTCTGCTTGGCTTG 2262  
Qy 170 TGAAGGAAGAGGAGGAGGACCACTGCTCCCGACAGACCCACCTGCAGGCA 219  
Db 2263 CGGGGCGGAAAGCGGAGGCTGAGCGCTCTCGGGCCCTCGGGGTGGGA 2312

Search completed: February 22, 2005, 00:06:48  
Job time : 166 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 23:10:41 ; Search time 500 Seconds  
(without alignments)  
7553.614 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696  
Perfect score: 639  
Sequence: 1 atcgctgcctccgcacg.....ataaccaggtgaccatcaag 639

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	261.8	41.0	497	9	US-09-864-761-156
C 2	239.8	37.5	243	9	US-09-864-761-16984
C 3	91.8	14.4	1128	18	US-10-363-345A-17159
C 4	91.8	14.4	1128	18	US-10-363-345A-17160
C 5	85.6	13.4	1128	18	US-10-363-345A-17157
C 6	85.6	13.4	1128	18	US-10-363-345A-17158
7	46.2	7.2	1770	17	US-10-425-114-25086
8	46.2	7.2	1770	18	US-10-425-115-165209
9	42.6	6.7	3018	9	US-09-802-127-6
10	42.6	6.7	3391	9	US-09-802-127-4
11	42.6	6.7	5968	9	US-09-764-868-1501
					Sequence 156, App
					Sequence 16984, A
					Sequence 17159, A
					Sequence 17160, A
					Sequence 17157, A
					Sequence 17158, A
					Sequence 25086, A
					Sequence 165209, A
					Sequence 6, Appli
					Sequence 4, Appli
					Sequence 1501, Ap

12	42.6	6.7	5968	14	US-10-103-313-619	Sequence 619, App
13	42.4	6.6	3368	18	US-10-437-963-13512	Sequence 13512, A
C 14	42.2	6.6	815	9	US-09-823-038A-9	Sequence 9, Appli
15	42	6.6	738	18	US-10-425-115-163054	Sequence 163054, A
16	41.6	6.5	149430	18	US-10-626-832-1	Sequence 1, Appli
C 17	41.6	6.5	149430	18	US-10-626-832-1	Sequence 1, Appli
C 18	41.6	6.5	150223	18	US-10-624-149A-1	Sequence 1, Appli
C 19	41.6	6.5	150223	18	US-10-624-149A-1	Sequence 1, Appli
20	41.4	6.5	537	18	US-10-437-963-66677	Sequence 66677, A
21	40.8	6.4	442	18	US-10-425-115-88722	Sequence 88722, A
C 22	40.8	6.4	1259	18	US-10-437-963-100621	Sequence 100621, A
C 23	40.8	6.4	1409	18	US-10-437-963-52890	Sequence 52890, A
C 24	40.6	6.4	101954	18	US-10-322-281-782	Sequence 782, App
25	40.6	6.4	200000	18	US-10-672-763A-31	Sequence 31, Appli
26	40.4	6.3	1732	18	US-10-437-963-25745	Sequence 25745, A
27	40.4	6.3	2000	17	US-10-260-238-23359	Sequence 2359, Ap
28	40.4	6.3	4426	17	US-10-366-547-46	Sequence 46, Appli
29	40.2	6.3	42339	13	US-10-087-192-991	Sequence 991, App
C 30	40	6.3	605	11	US-09-764-875-117	Sequence 117, App
C 31	39.8	6.2	345	18	US-10-425-115-94260	Sequence 94260, A
32	39.8	6.2	831	18	US-10-437-963-66678	Sequence 66678, A
33	39.8	6.2	981	17	US-10-425-114-22196	Sequence 22196, A
34	39.8	6.2	981	18	US-10-425-115-94549	Sequence 94549, A
35	39.8	6.2	15872	9	US-09-861-289-1	Sequence 1, Appli
36	39.8	6.2	15872	9	US-09-860-846-1	Sequence 1, Appli
37	39.8	6.2	15872	10	US-09-988-384B-1	Sequence 1, Appli
38	39.8	6.2	15872	10	US-09-836-821-1	Sequence 1, Appli
39	39.8	6.2	15872	16	US-10-371-889-44	Sequence 44, Appli
40	39.6	6.2	516	18	US-10-437-963-48622	Sequence 48622, A
41	39.6	6.2	1261	18	US-10-437-963-70724	Sequence 70724, A
42	39.4	6.2	653	13	US-10-027-632-132758	Sequence 132758, A
43	39.4	6.2	653	13	US-10-027-632-132759	Sequence 132759, A
44	39.4	6.2	653	13	US-10-027-632-132760	Sequence 132760, A
45	39.4	6.2	653	17	US-10-027-632-132758	Sequence 132758, A

ALIGNMENTS

RESULT 1

- US-09-864-761-156/c Application US/09864761
- Sequence 156, Application US/09864761
- Patent No. US20020048763A1
- GENERAL INFORMATION:
- APPLICANT: Penn, Sharon G.
- APPLICANT: Rank, David R.
- APPLICANT: Hanzel, David K.
- APPLICANT: Chen, Wensheng
- TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
- FILE REFERENCE: Aomic-X-1
- FILE REFERENCE: Aomic-X-1
- CURRENT APPLICATION NUMBER: US/09/864,761
- CURRENT FILING DATE: 2001-05-23
- PRIOR APPLICATION NUMBER: US 60/180,312
- PRIOR FILING DATE: 2000-02-04
- PRIOR APPLICATION NUMBER: US 60/207,456
- PRIOR FILING DATE: 2000-05-26
- PRIOR APPLICATION NUMBER: US 09/632,366
- PRIOR FILING DATE: 2000-08-03
- PRIOR APPLICATION NUMBER: GB 24263.6
- PRIOR FILING DATE: 2000-10-04
- PRIOR APPLICATION NUMBER: US 60/236,359
- PRIOR FILING DATE: 2000-09-27
- PRIOR APPLICATION NUMBER: PCT/US01/00666
- PRIOR FILING DATE: 2001-01-30
- PRIOR APPLICATION NUMBER: PCT/US01/00667
- PRIOR FILING DATE: 2001-01-30
- PRIOR APPLICATION NUMBER: PCT/US01/00664
- PRIOR FILING DATE: 2001-01-30
- PRIOR APPLICATION NUMBER: PCT/US01/00669
- PRIOR FILING DATE: 2001-01-30
- PRIOR APPLICATION NUMBER: PCT/US01/00665
- PRIOR FILING DATE: 2001-01-30



RESULT 4  
US-10-363-345A-17160/c  
; Sequence 17160, Application US/10363345A  
; Publication No. US20040234960A1

Query Match	13.4%	Score 85.6;	DB 18;	Length 1128;
Best Local Similarity	75.7%	Pred. No. 5.7e-14;		
Matches 106;	Conservative 0;	Mismatches 34;	Indels 0;	Gaps 0

  

Qy	1	ATGGTGCGGCTCGCGACCGAGCGGGCTCTCTCTGCGTGTGCTGCTGGCGGCGTG	60
Db	521	ATACGTACGCTCGCGACCGAACCGAACTCTCTCTACGTACTACTAACGACGCTA	462
Qy	61	CTGAGCGCGGCTAGGGCTCCCGTGAAGAGCGCGGCTCCGGGACACAGGCGCTGGG	120
Db	461	CTAAAAACGACGCTAAAACTCCCGTAAAAAAAACCGGACTCCCGGAACACGACCTAAA	402
Qy	121	AGCCTCACGAGGCTCGCAGA	140









APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 13512  
LENGTH: 3368  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_19539C.1  
US-10-437-963-13512

Query Match 6.6%; Score 42.4; DB 18; Length 3368;  
Best Local Similarity 50.0%; Pred. No. 0.062;  
Matches 106; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
QY 3 CGCTGCGCTCCGCGACCGAGCGGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62  
Db 3213 GCGCTGCGCCACCTCCGCGCGCGCTCGGCTCCACCGCTGCCCCCGCGCGCGCGCG 3154  
QY 63 GGAGCGCGGCTAGGCTCCCGTGAAGAGCGCGGCTCCGCGACCGACCGCTGGAG 122  
Db 3153 CGCCCTCGCGCGGAGGGTCCCCGAGGAGGCGCGCTCCGCTCGGCTCCCTCGG 3094  
QY 123 CCTACGAGGTCGCGAGAGGTCACGCTCCCGAGATCTAGGCTCTGAAGGAGAGGA 182  
Db 3093 CTTGCGCGCTCGCGCGCGCTTCTCTCTCGCTCCGAGCGCTCCGCGCGCT 3034  
QY 183 GGAGCACCACTGTCTCCCGAACCCACTGC 214  
Db 3033 CGTCCGCCACCTCGTCGCTACTACGACCGC 3002

RESULT 14  
US-09-823-038A-9  
Sequence 9, Application US/09823038A  
Patent No. US20020058335A1  
GENERAL INFORMATION:  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Abernethy, Nevin  
APPLICANT: Onrust, Rene  
APPLICANT: Kumble, Anand  
APPLICANT: Murison, Greg  
TITLE OF INVENTION: Compositions Isolated From Stromal Cells  
TITLE OF INVENTION: and Methods For Their Use  
FILE REFERENCE: 11000.1037c3  
CURRENT APPLICATION NUMBER: US/09/823,038A  
CURRENT FILING DATE: 2001-07-09  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 815  
TYPE: DNA  
ORGANISM: Mouse  
US-09-823-038A-9

Query Match 6.6%; Score 42.2; DB 9; Length 815;  
Best Local Similarity 70.9%; Pred. No. 0.065;  
Matches 56; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
QY 7 GCGCTCCGCGACCGAGCGGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 66  
Db 122 GCGCTGTCGCGCGCGCGCTGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTAGG 181  
QY 67 GCGGCGTAGGGCTCCCCG 85

Db 182 CGCGCGCAGGACCGACCG 200

## RESULT 15

US-10-425-115-163054  
Sequence 163054, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 163054  
LENGTH: 738  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_80279C.1  
US-10-425-115-163054

Query Match 6.6%; Score 42; DB 18; Length 738;  
Best Local Similarity 49.5%; Pred. No. 0.074;  
Matches 108; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 12 CCGGACCGAGCGGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAGGCGGC 71  
Db 90 CAGCGCGGACGCGCGACGCGCGCATGGCTCTACTAACGCGCGCCACCGCGCGCT 149  
QY 72 GCTAGGGTCCCGTGAAGAGCGCGGCTCCGCGGACCGGCTGCGGAGGAGGAGGAGGAGG 131  
Db 150 GGTGCTCTCGCTGCGCGCGGCTGCTGCGGCTGCTGCGGAGGAGGAGGAGGAGGAGG 209  
QY 132 GCTGCGAGGCTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAGGAGGAGGAGGAGGAGG 191  
Db 210 ATCAGCCCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269  
QY 192 ACTGCTCCCCAGAACCCACCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 229  
Db 270 GCTGAACCTGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307

Search completed: February 22, 2005, 01:26:30  
Job time : 501 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 22:00:30 ; Search time 3098 Seconds  
(without alignments)  
7851.216 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 639

Sequence: 1 atcgctgcgtccgcgacgcg.....ataaccaggtgaccatcaag 639

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gse1:\*  
9: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	637.4	99.7	790	5	BX336289
2	637.4	99.7	826	5	BX387234
3	637.4	99.7	866	5	BX365974
4	637.4	99.7	867	5	BX403263
5	637.4	99.7	868	5	BX365998
6	637.4	99.7	872	5	BX344975
7	637.4	99.7	900	5	BX398022
8	637.4	99.7	917	5	BX334584
9	637.4	99.7	936	5	BX357282
10	637.4	99.7	936	5	BX381620
11	637.4	99.7	947	5	BX359737
12	637.4	99.7	960	5	BX381477
13	637.4	99.7	968	5	BX356306
14	637.4	99.7	971	5	BX336623
15	637.4	99.7	973	5	BX379676
16	637.4	99.7	975	5	BX382082
17	637.4	99.7	978	5	BX336717
18	637.4	99.7	979	5	BX335053
19	637.4	99.7	979	5	BX360184
20	637.4	99.7	980	5	BX360878
21	637.4	99.7	983	1	AL554329
22	637.4	99.7	983	5	BX358077
23	637.4	99.7	983	5	BX361532
24	637.4	99.7	985	5	BX458619

#### ALIGNMENTS

RESULT 1  
BX336289  
LOCUS  
DEFINITION BX336289 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
clone CSODI027YA01 5-PRIME, mRNA sequence.  
ACCESSION BX336289  
VERSION BX336289.2 GI:46274766  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 2, 2003 this sequence version replaced gi:30341500.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r  
There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CSODI027AA01QP1&c=7348.r.

#### FEATURES

source  
1..790  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI027YA01"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

#### ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 790;  
Best Local Similarity 99.8%; Pred. No. 5.4e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

BX333696 BX333696  
CR749863 Homo sapi  
BX360064 BX360064  
BX357315 BX357315  
BX344130 BX344130  
BX345009 BX345009  
BX337459 BX337459  
BX387069 BX387069  
BX378610 BX378610  
BX343795 BX343795  
CNSLTI1BO  
BX248770 human ful  
BX324767 BX324767  
AL549276 AL549276  
BX387067 BX387067  
BX366007 BX366007  
BX397246 BX397246  
BX366004 BX366004  
BX335752 BX335752  
BX339514 BX339514  
BX359034 BX359034  
BX358526 BX358526

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?s=CS5AA0142D04RM1&c=7348.r>.

#### FEATURES

source  
1. 826  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DI034YP24"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

#### ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 826;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTGCTGTGCTGGCGGCGCTG	60
Db	5	ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTGCTGTGCTGGCGGCGCTG	64
Qy	61	CTGAGGCGGCGCTAGGGCTTCCCGCTGAAGAAGCCGCGCTCCGCGGACCAACGCTTGG	120
Db	65	CTGAGGCGGCGCTAGGGCTTCCCGCTGAAGAAGCCGCGCTCCGCGGACCAACGCTTGG	124
Qy	121	AGCCTACAGAGGTCGAGAGGTCCTAGCCTCCCGAGATCTAGGCTCTGAGGAGAG	180
Db	125	AGCCTACAGAGGTCGAGAGGTCCTAGCCTCCCGAGATCTAGGCTCTGAGGAGAG	184
Qy	181	GAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCCACCAACATGATGC	240
Db	185	GAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCCACCAACATGATGC	244
Qy	241	TGGAATCTCACTGAGCGAGAGCCATGCCCGAGGCAACACCCCTCCAGGACCCCA	300
Db	245	TGGAATCTCACTGAGCGAGAGCCATGCCCGAGGCAACACCCCTCCAGGACCCCA	304
Qy	301	GAGTTACTCGTTCGGCTGGAGTGCAGAGCTCCGGGATTTGGCAGCACAACTTG	360
Db	305	GAGTTACTCGTTCGGCTGGAGTGCAGAGCTCCGGGATTTGGCAGCACAACTTG	364
Qy	361	AGTACCCCTAACCTGATACCCAGCTTCCCGAGATCTAGGCTCTCAGGGAA	420
Db	365	AGTACCCCTAACCTGATACCCAGCTTCCCGAGATCTAGGCTCTCAGGGAA	424
Qy	421	GAGGAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCTACCAACATGGA	480
Db	425	GAGGAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCTACCAACATGGA	484
Qy	481	TGTTGAGTCTCACTGAGCGAGAGCCCTGACCCAGGAGATGACGCTCCAGGACC	540
Db	485	TGTTGAGTCTCACTGAGCGAGAGCCCTGACCCAGGAGATGACGCTCCAGGACC	544
Qy	541	CAGGAGGTTACTTCCCTTGTCTGAGAGTGCAGAGCTGCGAGATTTGGTCCACGCAAC	600
Db	545	CAGGAGGTTACTTCCCTTGTCTGAGAGTGCAGAGCTGCGAGATTTGGTCCACGCAAC	604
Qy	601	TTGAGTACCCCTAACCTGATACCCAGTGTACCATCAAG	639
Db	605	TTGAGTACCCCTAACCTGATACCCAGTGTACCATCAAG	643

#### RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

6365974 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI034YP24 5-PRIME, mRNA sequence.  
BX365974  
BX365974.2 GI:46288863

Qy	1	ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTGCTGTGCTGGCGGCGCTG	60
Db	15	ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTGCTGTGCTGGCGGCGCTG	74
Qy	61	CTGAGGCGGCGCTAGGGCTTCCCGCTGAAGAAGCCGCGCTCCGCGGACCAACGCTTGG	120
Db	75	CTGAGGCGGCGCTAGGGCTTCCCGCTGAAGAAGCCGCGCTCCGCGGACCAACGCTTGG	134
Qy	121	AGCCTACAGAGGTCGAGAGGTCCTAGCCTCCCGAGATCTAGGCTCTGAGGAGAG	180
Db	135	AGCCTACAGAGGTCGAGAGGTCCTAGCCTCCCGAGATCTAGGCTCTGAGGAGAG	194
Qy	181	GAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCCACCAACATGATGC	240
Db	195	GAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCCACCAACATGATGC	254
Qy	241	TGGAATCTCACTGAGCGAGAGCCATGCCCGAGGCAACACCCCTCCAGGACCCCA	300
Db	255	TGGAATCTCACTGAGCGAGAGCCATGCCCGAGGCAACACCCCTCCAGGACCCCA	314
Qy	301	GAGTTACTCGTTCGGCTGGAGTGCAGAGCTCCGGGATTTGGCAGCACAACTTG	360
Db	315	GAGTTACTCGTTCGGCTGGAGTGCAGAGCTCCGGGATTTGGCAGCACAACTTG	374
Qy	361	AGTACCCCTAACCTGATACCCAGCTTCCCGAGATCTAGGCTCTCAGGGAA	420
Db	375	AGTACCCCTAACCTGATACCCAGCTTCCCGAGATCTAGGCTCTCAGGGAA	434
Qy	421	GAGGAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCTACCAACATGGA	480
Db	435	GAGGAGGAGGACCACTGCTCCCGAGAACCCAGCTCGAGGAGAGCTACCAACATGGA	494
Qy	481	TGTTGAGTCTCACTGAGCGAGAGCCCTGACCCAGGAGATGCAACGCTCCAGGACC	540
Db	495	TGTTGAGTCTCACTGAGCGAGAGCCCTGACCCAGGAGATGCAACGCTCCAGGACC	554
Qy	541	CAGGAGGTTACTTCCCTTGTCTGAGAGTGCAGAGCTGCGAGATTTGGTCCACGCAAC	600
Db	555	CAGGAGGTTACTTCCCTTGTCTGAGAGTGCAGAGCTGCGAGATTTGGTCCACGCAAC	614
Qy	601	TTGAGTACCCCTAACCTGATACCCAGTGTACCATCAAG	639
Db	615	TTGAGTACCCCTAACCTGATACCCAGTGTACCATCAAG	653

RESULT 2  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX387234  
BX387234 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI034YP24 5-PRIME, mRNA sequence.  
BX387234  
BX387234.2 GI:46571918  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 8, 2003 this sequence version replaced gi:30447511.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: [seqrefgenoscope.cns.fr](mailto:seqrefgenoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
COMMENT 1 (bases 1 to 866)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30370960.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r.

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZF12\_AX21ZH4\_1&c=7348.r.

FEATURES  
source  
1..866  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="CS0D1031YB04"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and Ecor V sites of the pCMVSPORT 6 vector. Library was normalized."  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Query Match 99.7%; Score 637.4; DB 5; Length 866;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGTGGCTCCGGACCGAGCGGCTCTCTCTCTGCGTGTGCTGCGCGCGGTG 60  
Db 9 ATGGTGGCTCCGGACCGAGCGGCTCTCTCTCTGCGTGTGCTGCGCGCGGTG 68

Qy 61 CTGAGGCGCGCTAGGCGCTCCCGTGAAGAGCGCGCTCCGCGACACGCGCTGGG 120  
Db 69 CTGAGGCGCGCTAGGCGCTCCCGTGAAGAGCGCGCTCCGCGACACGCGCTGGG 128

Qy 121 AGCCTCAGCGGCTCGCAGAGGTCTCAGCCTCCCGAGATCTTAGGCTCTGAAGGAAG 180  
Db 129 AGCCTCAGCGGCTCGCAGAGGTCTCAGCCTCCCGAGATCTTAGGCTCTGAAGGAAG 188

Qy 181 GAGAGGACCACTGTCTCCAGAACCCACTGAGGAGGAGGACCAACATGATGC 240  
Db 189 GAGAGGACCACTGTCTCCAGAACCCACTGAGGAGGAGGACCAACATGATGC 248

Qy 241 TGGACTGTCTAGGCGAGGAGGATGACCCAGGACACACCCCTCCAGGACCCCA 300  
Db 249 TGGACTGTCTAGGCGAGGAGGATGACCCAGGACACACCCCTCCAGGACCCCA 308

Qy 301 GAGGTACTCCGTTGCGGCTGGAGCTGCAGAGGTGCGGGATTGGCCAGCACACCTTG 360  
Db 309 GAGGTACTCCGTTGCGGCTGGAGCTGCAGAGGTGCGGGATTGGCCAGCACACCTTG 368

Qy 361 AGTACCCCTAACCTGTATACCCAGCTTACGCTTCCCGAGATCTTAGGCTCTGAGGAA 420  
Db 369 AGTACCCCTAACCTGTATACCCAGCTTACGCTTCCCGAGATCTTAGGCTCTGAGGAA 428

Qy 421 GAGAGGAGGACGACTGTCTCCAGAACCCACTGAGGAGGAGGACTACACCATGGA 480  
Db 429 GAGAGGAGGACGACTGTCTCCAGAACCCACTGAGGAGGAGGACTACACCATGGA 488

Qy 481 TGTGGACTGTCTACTGAGCAGCAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACC 540  
Db 489 TGTGGACTGTCTACTGAGCAGCAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACC 548

Qy 541 CAGGAGGTACTCCCTTGTGCTGAGCTGCAGAGCTGCCAGAAATTGGTCCAGCAGACC 600  
Db 549 CAGGAGGTACTCCCTTGTGCTGAGCTGCAGAGCTGCCAGAAATTGGTCCAGCAGACC 608

Qy 601 TTGAGTACCCCTAACCCCTGATACACGAGTGACCATCAAG 639  
Db 609 TTGAGTACCCCTAACCCCTGATACACGAGTGACCATCAAG 647

RESULT:4  
BX403263  
LOCUS  
DEFINITION BX403263 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
Clone CS0D1075YH10 5-PRIME, mRNA sequence.  
ACCESSION BX403263  
VERSION BX403263.2 GI:46846816  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 867)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 13, 2003 this sequence version replaced gi:30615118.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r.

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS2BAX14ZE09\_AX24ZH3\_1&c=7348.r.

FEATURES  
source  
1..867  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="CS0D1075YH10"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and Ecor V sites of the pCMVSPORT 6 vector. Library was normalized."  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Query Match 99.7%; Score 637.4; DB 5; Length 867;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGTGGCTCCCGACCGAGCGGCTCTCTCTCTGCGTGTGCTGCGCGCGGTG 60  
Db 28 ATGGTGGCTCCCGACCGAGCGGCTCTCTCTCTGCGTGTGCTGCGCGCGGTG 87

Qy 61 CTGAGGCGCGCTAGGCGCTCCCGTGAAGAGCGCGGCTCCCGGAGCACAGCGCTGG 120  
Db 88 CTGAGGCGCGCTAGGCGCTCCCGTGAAGAGCGCGGCTCCCGGAGCACAGCGCTGG 147

Qy 121 AGCCTCAGGAGGCTCGCAGAGGTCTCAGCCTCCCGAGATCTTAGGCTCTGAAGGAAG 180  
Db 148 AGCCTCAGGAGGCTCGCAGAGGTCTCAGCCTCCCGAGATCTTAGGCTCTGAAGGAAG 207





Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see [http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZA01\\_AX19ZH11\\_l&c=7348.r](http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZA01_AX19ZH11_l&c=7348.r)

FEATURES  
source  
Location/Qualifiers  
1..872  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI005YP06"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
ORIGIN  
Query Match 99.7%; Score 637.4; DB 5; Length 872;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGGCTCCGGACCGCGGCTCTCTCTGCGTGTGCTGCGGCGCTG 60  
Db 13 ATGCGTGGCTCCCGACCGCGGCTCTCTCTGCGTGTGCTGCGGCGCTG 72  
Qy 61 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 120  
Db 73 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 132  
Qy 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTAGGCTCTGAAGGAAG 180  
Db 133 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTAGGCTCTGAAGGAAG 192  
Qy 181 GAGGAGGACCACTGCTCCCGAGACCACTGCGAGGAGGAGGACCAACATGGATGC 240  
Db 193 GAGGAGGACCACTGCTCCCGAGACCACTGCGAGGAGGAGGACCAACATGGATGC 252  
Qy 241 TGGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 253 TGGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312  
Qy 301 GAGGTTACTCCGTTGCGGCTGAGCTGCGAAGCTGCGGGAATGGCCAGCACAACTTG 360  
Db 313 GAGGTTACTCCGTTGCGGCTGAGCTGCGAAGCTGCGGGAATGGCCAGCACAACTTG 372  
Qy 361 AGTACCCTTAACCTGATACCGAGGCTTACGCTCTCCAGATCTAGGCTCTGAGGAA 420  
Db 373 AGTACCCTTAACCTGATACCGAGGCTTACGCTCTCCAGATCTAGGCTCTGAGGAA 432  
Qy 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
Db 433 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 492  
Qy 481 TGTGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
Db 493 TGTGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552  
Qy 541 CAGGAGGTTACTCCCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
Db 553 CAGGAGGTTACTCCCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 612  
Qy 601 TTGAGTACCCCTAACCTGATAACGAGTGACCATCAAG 639  
Db 613 TTGAGTACCCCTAACCTGATAACGAGTGACCATCAAG 651

RESULT 7  
BX398022  
LOCUS  
DEFINITION  
BX398022 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CSODI051YH09 5-PRIME, mRNA sequence.  
BX398022  
ACCESSION  
BX398022.2 GI:46876522  
VERSION  
BX398022  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 900)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
On May 13, 2003 this sequence version replaced gi:30621466.  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see [http://www.genoscope.cns.fr/cdna?s=CSODI051YH09\\_l&c=7348.r](http://www.genoscope.cns.fr/cdna?s=CSODI051YH09_l&c=7348.r).

FEATURES  
source  
Location/Qualifiers  
1..900  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI051YH09"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 900;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCGTGGCTCCCGACCGCGGCTCTCTCTGCGTGTGCTGCGGCGCTG 60  
Db 13 ATGCGTGGCTCCCGACCGCGGCTCTCTCTGCGTGTGCTGCGGCGCTG 72  
Qy 61 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 120  
Db 73 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 132  
Qy 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTAGGCTCTGAAGGAAG 180  
Db 133 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTAGGCTCTGAAGGAAG 192  
Qy 181 GAGGAGGACCACTGCTCCCGAGACCACTGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
Db 193 GAGGAGGACCACTGCTCCCGAGACCACTGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 252  
Qy 241 TGGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 253 TGGAGTGTCACTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312  
Qy 301 GAGGTTACTCCGTTGCGGCTGAGCTGCGAAGCTGCGGGAATGGCCAGCACAACTTG 360  
Db 313 GAGGTTACTCCGTTGCGGCTGAGCTGCGAAGCTGCGGGAATGGCCAGCACAACTTG 372





```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI023YN23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match      99.7%; Score 637.4; DB 5; Length 936;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1  ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 60
Db      9  ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 68
Qy     61  CTGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCTCCGCGACCAACGCCCTGGG 120
Db     69  CTGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCTCCGCGACCAACGCCCTGGG 128
Qy    121  AGCCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAAGAG 180
Db    129  AGCCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAAGAG 188
Qy    181  GAGGAGGACCACTGCTCCCGAGAACCCACTGCGAGGAGGACCAACATGGATGC 240
Db    189  GAGGAGGACCACTGCTCCCGAGAACCCACTGCGAGGAGGACCAACATGGATGC 248
Qy    241  TGGACTGTCTAGCGCAGCAGCCATGACCCCGAGCAACACACCCCTCCCGAGACCCA 300
Db    249  TGGACTGTCTAGCGCAGCAGCCATGACCCCGAGCAACACACCCCTCCCGAGACCCA 308
Qy    301  GAGGTTACTCGTTGCGGCTGAGCTGAGAGCTGCGGAGCTGCGGAGTTGGCGAGCAACCTTG 360
Db    309  GAGGTTACTCGTTGCGGCTGAGCTGAGAGCTGCGGAGCTGCGGAGTTGGCGAGCAACCTTG 368
Qy    361  AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGGAA 420
Db    369  AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGGAA 428
Qy    421  GAGGAGGAGGACGACTGCTCCCGAGAACCCACTGCGAGGAGGAGTACACCAATGGA 480
Db    429  GAGGAGGAGGACGACTGCTCCCGAGAACCCACTGCGAGGAGGAGTACACCAATGGA 488
Qy    481  TGTGGACTGTCTAGCGCAGCAGCCCTGACCCCGAGGAGTGCAGGCTCCCGAGACC 540
Db    489  TGTGGACTGTCTAGCGCAGCAGCCCTGACCCCGAGGAGTGCAGGCTCCCGAGACC 548
Qy    541  CAGGAGGTTACTCCCTTGTCTGCTGAGCTGAGAGCTGCCAGAGTTGGTCCAGCAACC 600
Db    549  CAGGAGGTTACTCCCTTGTCTGCTGAGCTGAGAGCTGCCAGAGTTGGTCCAGCAACC 608
Qy    601  TTGAGTACCCCTAACCTTGATACCCAGGAGTACCATCAAG 639
Db    609  TTGAGTACCCCTAACCTTGATACCCAGGTTACCATCAAG 647

RESULT 10
LOCUS   BX381620
DEFINITION BX381620 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI068YE01 5-PRIME, mRNA sequence.
ACCESSION BX381620
VERSION   BX381620.2 GI:46570758
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 936)
```

```
AUTHORS  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On May 8, 2003 this sequence version replaced gi:30443700.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqraf@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and EcoR V
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7348.r
There is a virtual cDNA representing this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/cdna?s=CS0DI068AC01Q1&c=7348.r.
FEATURES
Location/Qualifiers
1. 936
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI068YE01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match      99.7%; Score 637.4; DB 5; Length 936;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1  ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 60
Db     13  ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 72
Qy     61  CTGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCTCCGCGACCAACGCCCTGGG 120
Db     73  CTGAGGCGGCGCTAGGGCTCCCGTGAAGAGCGCGGCTCCGCGACCAACGCCCTGGG 132
Qy    121  AGCCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAGAGGAAGAG 180
Db    133  AGCCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAGAGGAAGAG 192
Qy    181  GAGGAGGACCACTGCTCCCGAGAACCCACTGCGAGGAGGACCAACCAATGGATGC 240
Db    193  GAGGAGGACCACTGCTCCCGAGAACCCACTGCGAGGAGGACCAACCAATGGATGC 252
Qy    241  TGGACTGTCTAGCGCAGCAGCCATGACCCCGAGCAACACCCCTCCCGAGACCCCA 300
Db    253  TGGACTGTCTAGCGCAGCAGCCATGACCCCGAGCAACACCCCTCCCGAGACCCCA 312
Qy    301  GAGGTTACTCGTTGCGGCTGAGCTGAGAGCTGCCGGGATTGGCGAGCAGCAACCTTG 360
Db    313  GAGGTTACTCGTTGCGGCTGAGCTGAGAGCTGCCGGGATTGGCGAGCAGCAACCTTG 372
Qy    361  AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGGAA 420
Db    373  AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGGAA 432
Qy    421  GAGGAGGAGGACGACTGCTCCCGAGAACCCACTGCGAGGAGGAGTACACCAATGGA 480
Db    433  GAGGAGGAGGACGACTGCTCCCGAGAACCCACTGCGAGGAGGAGTACACCAATGGA 492
Qy    481  TGTGGACTGTCTAGCGCAGCAGCCCTGACCCCGAGGAGTGCAGGCTCCCGAGACC 540
Db    493  TGTGGACTGTCTAGCGCAGCAGCCCTGACCCCGAGGAGTGCAGGCTCCCGAGACC 552
Qy    541  CAGGAGGTTACTCCCTTGTCTGCTGAGCTGAGAGCTGCCAGAGTTGGTCCAGCAACC 600
```



Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCTGCGCTCCGCGAGCGAGCGGGCTCTCTCTCGCTGCTGCTGCTGCGCGCGCTG 60  
Db 14 ATGCTGCGCTCCGCGAGCGAGCGGGCTCTCTCTCGCTGCTGCTGCTGCGCGCGCTG 73

Qy 61 CTGAGGCGCGCTAGGCGCTCCGCTGAGAGCGCGGCTCCGCGAGCGAGCGCGCTG 120  
Db 74 CTGAGGCGCGCTAGGCGCTCCGCTGAGAGCGCGGCTCCGCGAGCGAGCGCGCTG 133

Qy 121 AGCTCTCAGAGGCTCGAGAGCTCTCAGCTCCCGAGATCTAGGCTCTTGAAGAGAG 180  
Db 134 AGCTCTCAGAGGCTCGAGAGCTCTCAGCTCCCGAGATCTAGGCTCTTGAAGAGAG 193

Qy 181 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
Db 194 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 253

Qy 241 TGGAGTGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 254 TGGAGTGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 313

Qy 301 GAGGTTACTCTGCTGCGGCTGAGGCTGAGAGGCTGCGGATGCGGAGGAGGAGGAG 360  
Db 314 GAGGTTACTCTGCTGCGGCTGAGGCTGAGAGGCTGCGGATGCGGAGGAGGAGGAG 373

Qy 361 AGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 420  
Db 374 AGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 433

Qy 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
Db 434 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 493

Qy 481 TGTGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGAGGAGGAGGAGGAGGAG 540  
Db 494 TGTGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGAGGAGGAGGAGGAGGAG 553

Qy 541 CAGGAGGTTACTCTCTGCTGCTGAGGCTGAGAGGCTGAGAGGAGGAGGAGGAGGAG 600  
Db 554 CAGGAGGTTACTCTCTGCTGCTGAGGCTGAGAGGCTGAGAGGAGGAGGAGGAGGAG 613

## RESULT 13

BX356306

LOCUS

DEFINITION BX356306 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI010YD02 5-PRIME, mRNA sequence.

ACCESSION

BX356306

VERSION

BX356306.2

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (Bases 1 to 968)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 5, 2003 this sequence version replaced gi:30374043.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with NotI and cloned

into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?c=CSODI010YD01QPI&c=7348.r>.

## FEATURES

source

1..968

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODI010YD02"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with NotI and cloned into the NotI and EcoRV

sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 968;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGGCTCCGCGAGCGAGCGGGCTCTCTCTCGCTGCTGCTGCTGCGCGCGCTG 60  
Db 14 ATGCGTGGCTCCGCGAGCGAGCGGGCTCTCTCTCGCTGCTGCTGCTGCGCGCGCTG 73

Qy 61 CTGAGGCGCGCTAGGCGCTCCGCTGAGAGCGCGGCTCCGCGAGCGAGCGCGCTGG 120  
Db 74 CTGAGGCGCGCTAGGCGCTCCGCTGAGAGCGCGGCTCCGCGAGCGAGCGCGCTGG 133

Qy 121 AGCTCTCAGAGGCTCGAGAGGCTCTCAGCTCCCGAGATCTAGGCTCTTGAAGAGAG 180  
Db 134 AGCTCTCAGAGGCTCGAGAGGCTCTCAGCTCCCGAGATCTAGGCTCTTGAAGAGAG 193

Qy 181 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
Db 194 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 253

Qy 241 TGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 300  
Db 254 TGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 313

Qy 301 GAGTTTACTCGTTGCGGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 360  
Db 314 GAGTTTACTCGTTGCGGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 373

Qy 361 AGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 420  
Db 374 AGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 433

Qy 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
Db 434 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 493

Qy 481 TGTGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 540  
Db 494 TGTGGAGTGTACTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 553

Qy 541 CAGGAGGTTACTCTCTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 600  
Db 554 CAGGAGGTTACTCTCTGCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 613

Qy 601 TTGAGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 639  
Db 614 TTGAGTACCCCTAACCTGATACCGAGCTTACGCTCCCGAGATCTAGGCTCTTGAAGAG 652

## RESULT 14

BX336623

LOCUS

DEFINITION

BX336623 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CSODI030Y019 5-PRIME, mRNA sequence.

ACCESSION

BX336623

BX336623

971 bp mRNA linear EST 07-APR-2004



```
Qy 181 GAGGAGGACCACTGCTCTCCAGAAACCCACCTGAGGAGAGCCACACCAACATGGATGC 240
Db |||||
Qy 195 GAGGAGGACCACTGCTCTCCAGAAACCCACCTGAGGAGAGCCACACCAACATGGATGC 254
Db |||||
Qy 241 TGGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 300
Db |||||
Qy 255 TGGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 314
Db |||||
Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAAAGCTGCCGGGATTGGCCAGCACACACCTTG 360
Db |||||
Qy 315 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAAAGCTGCCGGGATTGGCCAAACACACCTTG 374
Db |||||
Qy 361 AGTACCCCTAACCTGTATACCCAGGCTTCAGGCTCCCCAGATCCCTAGGCTCTGAGGGAA 420
Db |||||
Qy 375 AGTACCCCTAACCTGTATACCCAGGCTTCAGGCTCCCCAGATCCCTAGGCTCTGAGGGAA 434
Db |||||
Qy 421 GAGGAGGAGGACGACTGCTCTCCAGAAACCCACCTGCAGGAGAGCTACACCAACATGGA 480
Db |||||
Qy 435 GAGGAGGAGGACGACTGCTCTCCAGAAACCCACCTGCAGGAGAGCTACACCAACATGGA 494
Db |||||
Qy 481 TGTGGACTGTCACTGAGCCAGAGCCCTGACCCAGGGAATGCCAGCCCTCCAGGACC 540
Db |||||
Qy 495 TGTGGACTGTCACTGAGCCAGAGCCCTGACCCAGGGAATGCCAGCCCTCCAGGACC 554
Db |||||
Qy 541 CAGGAGGTTACTCCCTTGCTGCTGAGCTGCAGAAAGCTGCAGAAATTGGTCCAGCAACC 600
Db |||||
Qy 555 CAGGAGGTTACTCCCTTGCTGCTGAGCTGCAGAAAGCTGCAGAAATTGGTCCAGCAACC 614
Db |||||
Qy 601 TTGAGTACCCCTAACCTGTATACCCAGGTTGACCATCAG 639
Db |||||
Qy 615 TTGAGTACCCCTAACCTGTATACCCAGGTTGACCATCAG 653
Db |||||
```

Search completed: February 22, 2005, 00:03:54  
Job time : 3102 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:41:21 ; Search time 169 Seconds  
(without alignments)  
427.954 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKPRLRGPRGFSUTRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	995	100.0	571	4 AAB49765	Aab49765 Human pro
2	995	100.0	571	4 AAB88393	Aab88393 Human mem
3	430	43.2	81	4 AAM13752	Aam13752 Peptide #
4	430	43.2	81	4 ABB32685	Abb32685 Peptide #
5	430	43.2	81	4 AAM26150	Aam26150 Peptide #
6	430	43.2	81	4 ABB27526	Abb27526 Human pep
7	430	43.2	81	4 ABB18176	Abb18176 Protein #
8	430	43.2	81	4 AAM65885	Aam65885 Human bon
9	430	43.2	81	4 AAM53507	Aam53507 Human bra
10	430	43.2	81	4 ABG47540	Abg47540 Human liv
11	430	43.2	81	4 AAM01497	Aam01497 Peptide #
12	430	43.2	81	5 ABG35520	Abg35520 Human pep
13	144	14.5	546	4 ABUS3252	Abu53252 Human tes
14	144	14.5	1270	8 ADK67911	Adk67911 Human ext
15	144	14.5	1299	4 AAM24322	Aam24322 Human EST
16	144	14.5	1311	8 ADK67912	Adk67912 Human ext
17	144	14.5	1320	7 ADK65819	Adk65819 Angiogene
18	144	14.5	1404	2 AAR26049	Aar26049 MSF precu
19	144	14.5	1404	4 AAB29773	Aab29773 Human meg
20	144	14.5	1404	4 AAB60568	Aab60568 Human meg
21	144	14.5	1404	7 ADK65839	Adk65839 Angiogene
22	144	14.5	1404	8 ADM98014	Adm98014 Human meg
23	144	14.5	1415	4 AAU32262	Aau32262 Novel hum
24	143.5	14.4	513	4 ABUS3254	Abu53254 Human tes
25	140.5	14.1	551	4 ABUS3253	Abu53253 Human tes

26	127	12.8	399	3 AAB43375	Aab43375 Human ORF
27	122	12.3	1171	4 ABB64387	Abb64387 Drosophil
28	118	11.9	345	8 ABO58798	AbO58798 Human gen
29	115.5	11.6	1518	6 ABJ18375	AbJ18375 Breast ep
30	114.5	11.5	235	6 ADA57458	Ada57458 Human sec
31	114.5	11.5	244	3 AAB38427	Aab38427 Fragment
32	114.5	11.5	268	7 ADD71122	Add71122 Human int
33	114.5	11.5	329	7 ADM04162	Adm04162 Human pro
34	114.5	11.5	406	4 AAU01623	Aau01623 Human sec
35	114.5	11.5	608	8 ADQ66076	Adq66076 Novel hum
36	114.5	11.5	681	4 ABG07912	Abg07912 Novel hum
37	114.5	11.5	726	4 AAB95371	Aab95371 Human pro
38	114.5	11.5	803	4 AAB94325	Aab94325 Human pro
39	114.5	11.5	803	4 AAG67424	Aag67424 Amino aci
40	114.5	11.5	803	5 ABB08164	Abb08164 Human cyt
41	114.5	11.5	803	5 ABG30717	Abg30717 Human Rho
42	114.5	11.5	803	5 ABB98248	Abb98248 Human Rho
43	114.5	11.5	803	8 ABM81720	Abm81720 Tumour-as
44	114.5	11.5	881	4 AAB90819	Aab90819 Human she
45	114.5	11.5	881	4 AAG65820	Aag65820 Human GTP

ALIGNMENTS

RESULT 1  
AAB49765  
ID AAB49765 standard; protein; 571 AA.  
XX  
AC AAB49765;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human proliferation differentiation factor amino acid sequence.  
XX  
KW Human; proliferation differentiation factor; haematopoietic function.  
XX  
OS Homo sapiens.  
XX  
PN W0200104312-A1.  
XX  
PD 18-JAN-2001.  
XX  
PF 06-JUL-2000; 2000WO-JP004514.  
XX  
PR 08-JUL-1999; 99JP-00194179.  
PR 18-OCT-1999; 99US-0159586P.  
XX  
(HELI-) HELIX RES INST.  
Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
WPI; 2001-138354/14.  
N-PSDB; AAF29348.  
Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cDNA library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.  
Claim 1; Page 42-44; 49pp; Japanese.  
This invention relates to polynucleotide sequence PSEC137, which encodes a human proliferation differentiation factor protein. Included in the invention is a vector containing the cDNA sequence, a transformant containing the vector, and a process for producing the protein. The protein together with its encoded DNA and antisense DNA are used in drugs to maintain the haematopoietic function. The present sequence represents the human proliferation differentiation factor protein amino acid sequence

SQ Sequence 571 AA;

Query Match 100.0%; Score 995; DB 4; Length 571;

Best Local Similarity 100.0%; Pred. No. 1.6e-75; Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPVKPRLRGPRGSLTRLAEVASDPDRPLKEEEAPLLPRTHLQAEPHQHCWTVTEP 60  
Db 27 LPVKPRLRGPRGSLTRLAEVASDPDRPLKEEEAPLLPRTHLQAEPHQHCWTVTEP 86

Qy 61 AAMTPGNTTPTPTPEVTPRLLEQLKPLGLASTTLLSTNPDTQASASDPDRPLREEEARL 120  
Db 87 AAMTPGNTTPTPTPEVTPRLLEQLKPLGLASTTLLSTNPDTQASASDPDRPLREEEARL 146

Qy 121 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLLLEQLKPLVHATLSTPNP 180  
Db 147 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLLLEQLKPLVHATLSTPNP 206

Qy 181 DNQVTIK 187  
Db 207 DNQVTIK 213

RESULT 2  
AAB88393  
ID AAB88393 standard; protein; 571 AA.  
AC AAB88393;  
DT 23-MAY-2001 (first entry)  
XX Human membrane or secretory protein clone PSEC0137.  
XX Human; secretory protein; membrane protein; vaccine; gene therapy;  
XX rheumatoid arthritis; diabetes.  
XX Homo sapiens.  
XX EP1067182-A2.  
XX 10-JAN-2001.  
XX 07-JUL-2000; 2000EP-00114090.  
XX 08-JUL-1999; 99JP-00194179.  
XX 11-JAN-2000; 2000JP-00118775.  
XX 02-MAY-2000; 2000JP-00183766.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
XX WPI; 2001-093989/11.  
XX N-PSDB; AAF93820.  
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
XX gene therapy or as candidate target molecules in drug development.  
XX Claim 1; SEQ ID NO 154; 609pp + Sequence Listing; English.  
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
XX which encode human secretory or membrane proteins represented by AAB88317  
XX - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
XX AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
XX invention. The invention also includes methods for the production of  
XX antibodies directed against the proteins, and cDNA sequences, which can  
XX be used in vaccines. The polynucleotide sequences can be used in gene  
XX therapy. The polynucleotide sequences and the proteins they encode may be  
XX used in the prevention, treatment and diagnosis of diseases associated  
XX with inappropriate secretory protein/membrane protein expression. The  
XX nucleic acids and complementary sequences may also be used as DNA probes  
XX in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
XX and quantitate the presence of similar nucleic acid sequences in samples.  
XX They may also be used to study the expression and function of secretory  
XX proteins/membrane polypeptides and their role in metabolism. The  
XX polypeptides may be used as antigens in the production of antibodies

CC against them and in assays to identify modulators (agonists and  
CC antagonists) of expression and activity. The antibodies and antagonists  
CC may also be used as therapeutic agents to down regulate expression and  
CC activity. The antibodies may also be used as diagnostic agents for  
CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
CC linked immunosorbant assay (ELISA). Examples of diseases which may be  
CC treated include rheumatoid arthritis and diabetes

XX  
SQ Sequence 571 AA;

Query Match 100.0%; Score 995; DB 4; Length 571;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPVKPRLRGPRGSLTRLAEVASDPDRPLKEEEAPLLPRTHLQAEPHQHCWTVTEP 60  
Db 27 LPVKPRLRGPRGSLTRLAEVASDPDRPLKEEEAPLLPRTHLQAEPHQHCWTVTEP 86

Qy 61 AAMTPGNTTPTPTPEVTPRLLEQLKPLGLASTTLLSTNPDTQASASDPDRPLREEEARL 120  
Db 87 AAMTPGNTTPTPTPEVTPRLLEQLKPLGLASTTLLSTNPDTQASASDPDRPLREEEARL 146

Qy 121 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLLLEQLKPLVHATLSTPNP 180  
Db 147 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLLLEQLKPLVHATLSTPNP 206

Qy 181 DNQVTIK 187  
Db 207 DNQVTIK 213

RESULT 3  
AAM13752  
ID AAM13752 standard; protein; 81 AA.  
AC AAM13752;  
DT 12-OCT-2001 (first entry)  
XX Peptide #186 encoded by probe for measuring cervical gene expression.  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX Homo sapiens.  
XX WO200157278-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US0000670.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
XX Claim 27; SEQ ID NO 18578; 487pp; English.  
XX The present invention relates to human single exon nucleic acid probes  
XX (SENP; see AAI10068-AAI28459). The present sequence is a peptide encoded



CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 22 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
Db 1 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60  
Qy 82 ELQKPLGLASTLSTPNPDQ 102  
Db 61 ELQKPLGLANTLSTPNPDQ 81

## RESULT 4

ID ABB32685 standard; peptide; 81 AA.

XX ABB32685;

XX 04-FEB-2002 (first entry)

XX Peptide #191 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.

XX Claim 27; SEQ ID NO 25320; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human fetal liver. The

XX present sequence is a peptide encoded by a single exon nucleic acid probe

XX of the invention. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 22 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
Db 1 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60

Qy 82 ELQKPLGLASTLSTPNPDQ 102

Db 61 ELQKPLGLANTLSTPNPDQ 81

## RESULT 5

AAM26150

ID AAM26150 standard; protein; 81 AA.

XX AAM26150;

XX 17-OCT-2001 (first entry)

XX Peptide #187 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 26419; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs:

XX see AA131315-AA157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders

XX Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 22 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
Db 1 VSASDPRLKKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60

Qy 82 ELQKPLGLASTLSTPNPDQ 102

```
Db      61 ELQKLPGLANTTLSTPNPDQ 81
RESULT 6
ABB27526
ID      ABB27526 standard; peptide; 81 AA.
XX
AC      ABB27526;
XX
DT      01-FEB-2002 (first entry)
XX
DE      Human peptide #177 encoded by breast cell single exon nucleic acid probe.
XX
KW      Human; microarray; single exon probe; gene expression; breast; disease;
XX      cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157271-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000662.
XX
PR      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00632366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-496933/54.
XX
PT      New spatially-addressable set of single exon nucleic acid probes, useful
XX      for measuring gene expression in sample derived from human breast,
XX      comprises number of single exon nucleic acid probes.
XX
PS      Claim 27; SEQ ID NO 10494; 327pp + Sequence Listing; English.
XX
CC      The invention relates to a spatially-addressable set of single exon
XX      nucleic acid probes for measuring gene expression in a sample derived
XX      from human breast and BT 474 cells. The method involves contacting the
XX      probes with a collection of detectably labelled nucleic acids derived
XX      from mRNA of human breast, and then measuring the label bound to each
XX      probe of the microarray. The probes are useful for verifying the
XX      expression of regions of genomic DNA predicted to encode proteins. They
XX      are useful for gene discovery, and for determining predisposition and/or
XX      prognosing breast disease. Gene expression analysis is useful for
XX      assessing the toxicity of chemical agents on cells. The microarray of
XX      this invention presents a far greater diversity of probes for measuring
XX      gene expression, with far less bias than expressed sequence tag
XX      microarrays. The method is suitable for rapid production of functional
XX      information from genomic sequence. The present sequence is a peptide
XX      encoded by a single exon nucleic acid probe of the invention. Note: The
XX      sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 81 AA;
Query Match      43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. NO. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      22 VSASDPRLKKEEAPLLRTHLOAEPHOGCWTTEPAAMTGCNTTTPRTPEVTPLRL 81
DB      1 VSASDPRLKKEEAPLLRTHLOAEPHOGCWTTEPAAMTGCNTTTPRTPEVTPLRL 60
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
```

```
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
RESULT 7
ABB18176
ID      ABB18176 standard; protein; 81 AA.
XX
AC      ABB18176;
XX
DT      23-JAN-2002 (first entry)
XX
DE      Protein #175 encoded by probe for measuring heart cell gene expression.
XX
KW      Human; gene expression; heart; microarray; vascular system;
XX      cardiovascular disease; hypertension; cardiac arrhythmia;
XX      congenital heart disease.
XX
OS      Homo sapiens.
XX
PN      WO200157274-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000666.
XX
PR      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00632366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488899/53.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
XX      hearts.
XX
PS      Claim 15; SEQ ID NO 19946; 530pp; English.
XX
CC      The present invention relates to single exon nucleic acid probes for
XX      measuring human gene expression in a sample derived from human heart (see
XX      ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX      probe. The probes may be used for predicting, measuring and displaying
XX      gene expression in samples derived from the human heart via microarrays.
XX      By measuring gene expression, the probes are useful for predicting,
XX      diagnosing, grading, staging, monitoring and prognosing diseases of the
XX      human heart and vascular system e.g. cardiovascular disease
XX      hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX      sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 81 AA;
Query Match      43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. NO. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      22 VSASDPRLKKEEAPLLRTHLOAEPHOGCWTTEPAAMTGCNTTTPRTPEVTPLRL 81
DB      1 VSASDPRLKKEEAPLLRTHLOAEPHOGCWTTEPAAMTGCNTTTPRTPEVTPLRL 60
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
```

```
RESULT 8
AAM65885
ID AAM65885 standard; protein; 81 AA.
XX AC AAM65885;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26191.
XX KW Human: bone marrow expressed probe encoded protein analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN W0200157276-A2.
XX PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 26191; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX SQ Sequence 81 AA;
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 22 VSASDPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
Qy 82 ELQKLPGLASTTLSTPNPDQ 102
Db 61 ELQKLPGLASTTLSTPNPDQ 81
RESULT 9
AAM53507
ID AAM53507 standard; protein; 81 AA.
XX AC AAM53507;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 25612.
XX KW Human: brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN W0200157275-A2.
XX PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 25612; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX SQ Sequence 81 AA;
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 22 VSASDPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
Qy 82 ELQKLPGLASTTLSTPNPDQ 102
Db 61 ELQKLPGLASTTLSTPNPDQ 81
RESULT 10
ABG47540
ID ABG47540 standard; peptide; 81 AA.
XX AC ABG47540;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID No 26188.
XX KW Human: liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX PN W0200157273-A2.
XX PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000664.
```

```

XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 26188; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 81 AA;
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
QY 82 ELQKLPLGLASTTLSTPNPDQ 102
Db 61 ELQKLPLGLANTTLSTPNPDQ 81
RESULT 11
AA001497
ID AA001497 standard; protein; 81 AA.
XX AA001497;
XX 09-OCT-2001 (first entry)
XX Peptide #179 encoded by probe for measuring human breast gene expression.
XX Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
XX WO200157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001; 2001WO-US0000661.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.

```

```

PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX Claim 27; SEQ ID NO 10237; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAI00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative,
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 81 AA;
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
QY 82 ELQKLPLGLASTTLSTPNPDQ 102
Db 61 ELQKLPLGLANTTLSTPNPDQ 81
RESULT 12
ABG35520
ID ABG35520 standard; peptide; 81 AA.
XX AC ABG35520;
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 25185.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX OS
XX WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US0000665.

```

XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 25185; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridize at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridization of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridization to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 81 AA;  
SQ  
Query Match 43.2%; Score 430; DB 5; Length 81;  
Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
Matches 75; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 22 VSASDPRLKEEAEAPLLRTHLQAEPHQHCWTVTEPAAMTNGNTPTTPTTEVPLRL 81  
DB 1 VSASDPRLKEEAEAPLLRTHLQAEPHQHCWTVTEPAAMTNGNTPTTPTTEVPLRL 60  
QY 82 ELOKPLGLASTTLSTPNPDQ 102  
DB 61 ELOKPLGLASTTLSTPNPDQ 81

RESULT 13  
ABUS3252  
ID ABUS3252 standard; protein; 546 AA.  
XX  
AC ABUS3252;  
XX  
DT 14-APR-2003 (first entry)  
XX  
DE Human testes-derived DKFZphtes3\_4019 homologue #1.  
XX  
KW Human; gene therapy; vaccine; disease treatment; detection.  
XX  
OS Homo sapiens.  
XX  
PN WO200112659-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-IB001496.  
XX  
PR 18-AUG-1999; 99US-0149499P.  
PR 28-SEP-1999; 99US-0156503P.  
XX  
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX  
PI Wiemann S;  
XX  
WPI; 2001-327840/34.  
XX  
Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX  
PS Example III; Page 892; 1095pp; English.  
XX  
CC This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
SQ Sequence 546 AA;  
Query Match 14.5%; Score 144; DB 4; Length 546;  
Best Local Similarity 28.4%; Pred. No. 0.001;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASDPRLKEEAEAPLLRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
DB 68 TTKKEPAPTPKPEAPTPPKKAPAPTPKPEAPTPKPEAPTPKAPAPTPKAPAPTPK 127  
QY 72 RTPVTPRLRLQKPLGLASTTLSTPNPDQASASP-----DPRPLREEEAEAPLLRTHLQ 127  
DB 128 ETAPTPKLTPTTPEKLAFTTPEKAPTPPELAPTTPEPTTPEPTTPEAPTPKAAAP 187  
QY 128 AELHQHCWTVTEPAALT---PGNATPPTTPTTEVPLRLLELOKPLVHATLSTPNP 180  
DB 188 NTPKEPAPTPKPEAPTPKPEAPTPKAPAPTPKAPAPTPKAPAPTPKAPAPTPK 243  
RESULT 14  
ADK67911  
ID ADK67911 standard; protein; 1270 AA.  
XX  
AC ADK67911;  
XX  
DT 06-MAY-2004 (first entry)  
XX

## Human extracellular messenger (EXMES) polypeptide.

Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;  
antiallergic; anti-inflammatory; antidiabetic; neuroprotective;  
muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiporiatic;  
virucide; fungicide; antiparasitic; protozoacide; antihelminthic;  
cytostatic; gene therapy.

Homo sapiens.

Key Location/Qualifiers  
Peptide 1..29  
/label= Signal\_peptide  
/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29  
according to identification method"

WO2004013292-A2.

12-FEB-2004.

30-JUL-2003; 2003WO-US024084.

02-AUG-2002; 2002US-0400810P.

19-SEP-2002; 2002US-0412197P.

04-OCT-2002; 2002US-0416004P.

08-NOV-2002; 2002US-0424862P.

(INCY-) INCYTE CORP.

Elliot VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

Lee SY;

WPI; 2004-157116/15.

N-PSDB; ADK67916.

New extracellular messengers and nucleic acids, useful for diagnosing,  
treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes  
mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or  
autoimmune thyroiditis.

Claim 59; SEQ ID NO 4; 165pp; English.

The present sequence is that of novel human extracellular messenger  
(EXMES) Incyte ID NO: 7513017CD1 polypeptide. The protein shows homology  
to human megakaryocyte stimulating factor. The invention provides EXMES  
polynucleotides and polypeptides, as well as expression vectors, host  
cells, antibodies, agonists and antagonists, and methods for diagnosing,  
treating or preventing disorders associated with aberrant expression of  
EXMES, especially autoimmune and inflammatory disorders, cell  
proliferative disorders and endocrine disorders, e.g. adult respiratory  
distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's  
disease, diabetes mellitus, myasthenia gravis, osteoarthritis,  
osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,  
parasitic, protozoal or helminthic infections, cancers, autoimmune  
thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.  
Embodiments also provide methods for using the purified EXMES and/or  
their encoding polynucleotides for facilitating the drug discovery  
process, including determining of efficacy, dosage, toxicity and  
pharmacology, and for investigating the pathogenesis of diseases and  
medical conditions.

Sequence 1270 AA;  
Query Match 14.5%; Score 144; DB 8; Length 1270;  
Best Local Similarity 28.4%; Pred. No. 0.0026;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLPLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 421 TTKEPAPPTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 480  
QY 72 RTPVTPRLRLQKPLGLASTTSLTPNDTQASAP-----DPRPLREERARLLPRTHLQ 127

Db 481 ETAPTTPKLTPPTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTEEPAPTTPKAAAP 540  
QY 128 AELHQHCWTVTEPAALT---PGNATPRTQVTPVTPLELLELQKLPELVHATLSTNP 180  
Db 541 NTPKEPAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 596

## RESULT 15

AAM24322

ID AAM24322 standard; protein; 1299 AA.

XX AAM24322;

AC AAM24322;

DT 12-OCT-2001 (first entry)

DE Human EST encoded protein SEQ ID NO: 1847.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;  
KW gene therapy; nutrition.

OS Homo sapiens.

XX WO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002687.

XX 25-JAN-2000; 2000US-00491404.

XX 17-JUL-2000; 2000US-00617746.

XX 03-AUG-2000; 2000US-00631451.

XX 15-SEP-2000; 2000US-00663870.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

XX N-PSDB; AAH98981.

Isolated polypeptide for treatment of diseases, diagnostics, raising  
antibodies and research use.

PS Claim 20; Page 1198-1201; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention

SQ Sequence 1299 AA;

Query Match 14.5%; Score 144; DB 4; Length 1299;

Best Local Similarity 28.4%; Pred. No. 0.0027;

Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLPLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71

Db 555 TTKEPAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 614

QY 72 RTPVTPRLRLQKPLGLASTTSLTPNDTQASAP-----DPRPLREERARLLPRTHLQ 127

Db 615 ETAPTTPKLTPPTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTEEPAPTTPKAAAP 674

QY 128 AELHQHCWTVTEPAALT---PGNATPRTQVTPVTPLELLELQKLPELVHATLSTNP 180

Db 675 NTPKEPAPPTPKEPAPPTPKETAPTTKGTAPTTLKEBPAPTTPKKPAP 730

Search completed: February 18, 2005, 23:56:01  
Job time : 172 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005. Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:50:51 ; Search time 43 Seconds  
(without alignments)  
324.637 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKKPLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	14.5	941	4	US-07-757-022B-14
2	144	14.5	1022	4	US-07-757-022B-84
3	144	14.5	1038	4	US-07-757-022B-74
4	144	14.5	1049	4	US-07-757-022B-58
5	144	14.5	1140	4	US-07-757-022B-104
6	144	14.5	1270	4	US-07-757-022B-44
7	144	14.5	1311	4	US-07-757-022B-42
8	144	14.5	1313	4	US-07-757-022B-142
9	144	14.5	1314	4	US-07-757-022B-50
10	144	14.5	1320	4	US-07-757-022B-46
11	144	14.5	1320	4	US-07-757-022B-60
12	144	14.5	1320	4	US-10-164-595-58
13	144	14.5	1354	4	US-07-757-022B-48
14	144	14.5	1361	4	US-07-757-022B-40
15	144	14.5	1363	4	US-07-757-022B-52
16	144	14.5	1404	4	US-07-757-022B-2
17	144	14.5	1404	4	US-07-757-022B-62
18	144	14.5	1404	4	US-10-164-595-78
19	144	14.5	1404	4	US-09-298-970A-1
20	144	14.5	1411	4	US-09-949-016-10827
21	113	11.4	5179	4	US-09-538-092-1258
22	112	11.3	183	6	5168049-3
23	112	11.3	183	6	5168049-3
24	111.5	11.2	267	3	US-08-818-112-142
25	111.5	11.2	267	3	US-08-818-111-137
26	111.5	11.2	267	3	US-09-056-556-142
27	111.5	11.2	267	4	US-09-072-596-137

Sequence 142, App  
Sequence 945, App  
Sequence 17307, A  
Sequence 4, Appli  
Sequence 9700, Ap  
Sequence 8969, Ap  
Sequence 736, App  
Sequence 6668, Ap  
Sequence 2, Appli  
Sequence 143, App  
Sequence 254, App  
Sequence 34, Appl  
Sequence 110, App  
Sequence 4, Appli  
Sequence 90, Appl  
Sequence 5258, Ap

28 111.5 11.2 267 4 US-09-072-967-142  
29 108.5 10.9 705 4 US-09-538-092-945  
30 108.5 10.9 827 4 US-09-248-796A-17307  
31 107.5 10.8 696 3 US-08-906-865-4  
32 107.5 10.8 696 4 US-09-129-668-4  
33 107.5 10.8 713 4 US-09-949-016-9700  
34 106 10.7 288 4 US-09-949-016-8969  
35 105.5 10.6 1530 4 US-09-976-594-736  
36 105.5 10.6 1530 4 US-09-949-016-6668  
37 105 10.6 486 1 US-08-450-360-2  
38 105 10.6 699 4 US-10-237-551-143  
39 105 10.6 699 4 US-10-237-551-254  
40 105 10.6 1215 4 US-10-029-180-34  
41 105 10.6 1244 4 US-10-029-180-110  
42 104.5 10.5 447 1 US-08-450-360-4  
43 103.5 10.4 258 3 US-08-961-083-90  
44 103.5 10.4 258 4 US-09-536-784-90  
45 103.5 10.4 305 4 US-09-107-532A-5258

#### ALIGNMENTS

#### RESULT 1

US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 941 amino acids  
; TYPE: AMINO ACID

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-14

Query Match      14.5%; Score 144; DB 4; Length 941;
Best Local Similarity 28.4%; Pred. No. 2.4e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 356 TTTKEPAPTTPKPAPTTPKKPAPTTPKEPAPTTPKKPAPTTPKPAPTAPKEPAPTTPK 415
QY 72 RTPEVTPLRLLEQLKPLGLASTTSLPNPDTOASASP-----DPRPLREEEEARLLPRTHLQ 127
Db 416 ETAPTTPKLPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPPEAPTTPKAAP 475
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELOKLPELVHATILSTNP 180
Db 476 NTPKEPAPTTPKPAPTTPKKPAPTTPKETAPTTCKGTAPTTLKEPAPTTPKKPA 531

RESULT 2
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-74

Query Match      14.5%; Score 144; DB 4; Length 1022;
Best Local Similarity 28.4%; Pred. No. 2.7e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 437 TTTKEPAPTTPKPAPTTPKKPAPTTPKEPAPTTPKKPAPTTPKPAPTAPKEPAPTTPK 496
QY 72 RTPEVTPLRLLEQLKPLGLASTTSLPNPDTOASASP-----DPRPLREEEEARLLPRTHLQ 127
Db 497 ETAPTTPKLPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPPEAPTTPKAAP 556
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELOKLPELVHATILSTNP 180
Db 557 NTPKEPAPTTPKPAPTTPKKPAPTTPKETAPTTCKGTAPTTLKEPAPTTPKKPA 612

RESULT 3
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-74
```

Query Match 14.5%; Score 144; DB 4; Length 1038;  
Best Local Similarity 28.4%; Pred. No. 2.8e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 421 TTTKEPAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPKAPTAKEPAPTTPK 480  
Qy 72 RTPEVTPURLLOKPLGLASTTLPNDTQASAP-----DPRPLREEEEARLLPRTHLQ 127  
Db 481 ETAPTTPKLTPTTPEKLAAPTTPKAPTTPEELAPTTPEEPTTTPPEAPTTPKAAAP 540  
Qy 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELLOKPLDELVHATLSTNP 180  
Db 541 NTPKEPAPTTKPEAPTTKPKAPTTKPKAPTTKPKAPTTKPKAP 596

RESULT 4  
US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-58

Query Match 14.5%; Score 144; DB 4; Length 1049;  
Best Local Similarity 28.4%; Pred. No. 2.8e-05;

Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 464 TTTKEPAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPKAPTAKEPAPTTPK 523  
Qy 72 RTPEVTPURLLOKPLGLASTTLPNDTQASAP-----DPRPLREEEEARLLPRTHLQ 127  
Db 524 ETAPTTPKLTPTTPEKLAAPTTPKAPTTPEELAPTTPEEPTTTPPEAPTTPKAAAP 583  
Qy 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELLOKPLDELVHATLSTNP 180  
Db 584 NTPKEPAPTTKPEAPTTKPKAPTTKPKAPTTKPKAPTTKPKAP 639

RESULT 5  
US-07-757-022B-104  
; Sequence 104, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 104:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1140 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-104

Query Match 14.5%; Score 144; DB 4; Length 1140;  
Best Local Similarity 28.4%; Pred. No. 3.1e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;



Qy 72 RTEVTPLRLLEQLKPLGLASTLSTPNPDTOASASP-----DPRPLREEEEARLLPRTHLQ 127  
Db 522 ETAPTPKLLPTTPEKLAFTTPPEELAPTTPEEPAPTTPKKAAP 581  
Qy 128 AELHQHGWTVEPAALT---PGNATPRTQEVTPLLLEQLKPLVHATLSTNP 180  
Db 582 NTPKEPAPTTPEKPAFTTPKPAFTTPKETAPTTKGTAPTTLKEPAPTTPKKAP 637

RESULT 8  
US-07-757-022B-142  
; Sequence 142, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 142:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-142

Query Match 14.5%; Score 144; DB 4; Length 1313;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASDPPLRPLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 464 TTTKEPAPTTPEKPAFTTPKPAFTTPKETAPTTKGTAPTTLKEPAPTTPKKAP 523  
Qy 72 RTEVTPLRLLEQLKPLGLASTLSTPNPDTOASASP-----DPRPLREEEEARLLPRTHLQ 127

Db 524 ETAPTPKLLPTTPEKLAFTTPPEELAPTTPEEPAPTTPKKAAP 583  
Qy 128 AELHQHGWTVEPAALT---PGNATPRTQEVTPLLLEQLKPLVHATLSTNP 180  
Db 584 NTPKEPAPTTPEKPAFTTPKPAFTTPKETAPTTKGTAPTTLKEPAPTTPKKAP 639

RESULT 9  
US-07-757-022B-50  
; Sequence 50, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1314 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-50

Query Match 14.5%; Score 144; DB 4; Length 1314;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASDPPLRPLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 465 TTTKEPAPTTPEKPAFTTPKPAFTTPKETAPTTKGTAPTTLKEPAPTTPKKAP 524  
Qy 72 RTEVTPLRLLEQLKPLGLASTLSTPNPDTOASASP-----DPRPLREEEEARLLPRTHLQ 127  
Db 525 ETAPTPKLLPTTPEKLAFTTPPEELAPTTPEEPAPTTPKKAAP 584

QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTLPLLELQKLPELVHATLSTNP 180  
Db 585 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKPAP 640

RESULT 10  
US-07-757-022B-46  
; Sequence 46, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757.022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-5851  
; TELEFAX: (617)876-1170  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-46

Query Match 14.5%; Score 144; DB 4; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASPPRLKEEERAPLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 471 TTTKEPATTTPKEPATTTPKKPATTTPKEPATTTPKKPATTTPKKPATTTPK 530  
QY 72 RPEVTPRLLEQLKPLGLASTTLSTNPDTQASAP----DPRPLREEEERALLRTHLQ 127  
Db 531 ETAPTTPKKLTTPTEKLAFTTPEKAPATTPELAFTTPEEFTTTPPEPATTTPKAAAP 590  
QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTLPLLELQKLPELVHATLSTNP 180

Db 591 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKPAP 646

RESULT 11  
US-07-757-022B-60  
; Sequence 60, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757.022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-60

Query Match 14.5%; Score 144; DB 4; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASPPRLKEEERAPLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 471 TTTKEPATTTPKEPATTTPKKPATTTPKEPATTTPKKPATTTPKKPATTTPK 530  
QY 72 RPEVTPRLLEQLKPLGLASTTLSTNPDTQASAP----DPRPLREEEERALLRTHLQ 127  
Db 531 ETAPTTPKKLTTPTEKLAFTTPEKAPATTPELAFTTPEEFTTTPPEPATTTPKAAAP 590  
QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTLPLLELQKLPELVHATLSTNP 180  
Db 591 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKPAP 646



ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1361 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-40

Query Match 14.5%; Score 144; DB 4; Length 1361;  
Best Local Similarity 28.4%; Pred. No. 3.9e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASPPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
DB 512 TTTKEPATTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTAPKEPATTTPK 571  
QY 72 RTPEVTPRLRLQKLPGLASTTLSTPNPDTOASAP-----DPRPLREEEERALLPRTHLQ 127  
DB 572 ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPPEPAPTTPKAAAP 631  
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPLLLELQKLPELVHATLSTNP 180  
DB 632 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAAPTTLKEPATTTPKKPAP 687

## RESULT 15

US-07-757-022B-52  
Sequence 52, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match 14.5%; Score 144; DB 4; Length 1363;  
Best Local Similarity 28.4%; Pred. No. 3.9e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASPPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
DB 514 TTTKEPATTTPKEPATTTPKKPAPTTPKEPATTTPKKPAPTAPKEPATTTPK 573  
QY 72 RTPEVTPRLRLQKLPGLASTTLSTPNPDTOASAP-----DPRPLREEEERALLPRTHLQ 127  
DB 574 ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPPEPAPTTPKAAAP 633  
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPLLLELQKLPELVHATLSTNP 180  
DB 634 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAAPTTLKEPATTTPKKPAP 689

Search completed: February 19, 2005, 00:00:40  
Job time : 48 secs





Db 87 ERGAGAGKPLDTCRCVPTEGSTVOLIAFNAADVHSHGDYDSQTCIRVSASPDPRPKZE 146  
Qy 34 EEEAPLLPRTHLOAEPHOGHCWTVTEPAAMTPGNTTTPRTPEVTPLRLLEQLKPLGLASTT 93  
Db 147 EEEAPLLPRTHLOAEPHOGHCWTVTEPAAMTPGNTTTPRTPEVTPLRLLEQLKPLGLANTT 206  
Qy 94 LSTNPDTQASASDPDRPLRREEEARLLPRTHLOAELHOGHCWTVTEPAALTPGNATPPR 153  
Db 207 LSTNPDTQASASDPDRPLRREEEARLLPRTHLOAELHOGHCWTVTEPAALTPGNATPPR 266  
Qy 154 TOEVTPLLEQLKPLPELVHATLSTPNPDNQVTIK 187  
Db 267 TOEVTPLLEQLKPLPELVHATLSTPNPDNQVTIK 300

RESULT 2  
US-10-074-978A-403  
; Sequence 403, Application US/10074978A  
; Publication No. US20040010119A1  
; GENERAL INFORMATION:  
; APPLICANT: Leite, Mario  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Guo, Xiaojia (Sasha)  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Li, Li  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Casman, Stacie  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Blalock, Angela  
; APPLICANT: Ballinger, Robert  
; APPLICANT: Vernet, Corine  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Mezes, Peter S  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Heyes, Melvin P  
; APPLICANT: Herrman, John  
; APPLICANT: Pena, Carol E A  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Moore, No. US20040010119A1lle  
; APPLICANT: Sheno, Suresh  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Gunther, Erik  
; APPLICANT: Stone, Dave  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Feyman, John  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-269  
; CURRENT APPLICATION NUMBER: US/10/074,978A  
; CURRENT FILING DATE: 2003-01-07  
; PRIOR APPLICATION NUMBER: 60/268,221  
; PRIOR FILING DATE: 2001-02-12  
; PRIOR APPLICATION NUMBER: 60/335,109  
; PRIOR FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: 60/312,284  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/268,496  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: 60/276,703  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/330,293  
; PRIOR FILING DATE: 2001-10-18  
; PRIOR APPLICATION NUMBER: 60/322,127  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: 60/280,899  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/310,797

; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/268,646  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 547  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 403  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-074-978A-403  
  
Query Match 88.3%; Score 879; DB 15; Length 538;  
Best Local Similarity 98.8%; Pred. No. 4.1e-60;  
Matches 164; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 22 VSASPDPRPLKREBEAPLLPRTHLOAEPHOGHCWTVTEPAAMTPGNTTTPRTPEVTPLRL 81  
Db 15 VSASPDPRPLKREBEAPLLPRTHLOAEPHOGHCWTVTEPAAMTPGNTTTPRTPEVTPLRL 74  
  
Qy 82 ELQKLPLGLASTTLPSTNPDTQASASDPDRPLRREEEARLLPRTHLOAELHOGHCWTVTEP 141  
Db 75 ELQKLPLGLANTTLPSTNPDTQASASDPDRPLRREEEARLLPRTHLOAELHOGHCWTVTEP 134  
  
Qy 142 AALTPGNATPPRTQEVTPPLLEQLKPLPELVHATLSTPNPDNQVTIK 187  
Db 135 AALTPGNATPPRTQEVTPPLLEQLKPLPELVHATLSTPNPDNQVTIK 180  
  
RESULT 3  
US-09-864-761-33474  
; Sequence 33474, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/532,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 33474  
;; LENGTH: 81  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AF11168.2  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.85  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92  
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.82  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.75  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.92  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE281519.1, EVALUATE 1.30e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE731965.1, EVALUATE 2.00e-31  
;; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATE 5.30e+00  
US-09-864-761-33474

Query Match 43.2%; Score 430; DB 9; Length 81;  
Best Local Similarity 97.5%; Pred. No. 3.4e-26;  
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 22 VSASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMTGGNTTTPRTPEVTPLRL 81  
Db 1 VSASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMTGGNTTTPRTPEVTPLRL 60  
Qy 82 ELQKLPLGLASTLTSTPNPDQ 102  
Db 61 ELQKLPLGLASTLTSTPNPDQ 81

RESULT 4  
US-10-124-557-14  
; Sequence 14, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Bridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

;; APPLICATION NUMBER: US 07/546,114  
;; FILING DATE: 29-JUN-1990  
;; APPLICATION NUMBER: US 07/457,196  
;; FILING DATE: 29-DEC-1989  
;; APPLICATION NUMBER: US 07/390,901  
;; FILING DATE: 08-AUG-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Cserr, Luann  
;; REGISTRATION NUMBER: 31,822  
;; REFERENCE/DOCKET NUMBER: GI 5190  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)876-1170  
;; TELEFAX: (617)876-5851  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 941 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-124-557-14  
Query Match 14.5%; Score 144; DB 13; Length 941;  
Best Local Similarity 28.4%; Pred. No. 0.0083;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
Qy 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 356 TTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 415  
Qy 72 RTPEVTPLRLQKLPLGLASTLTSTPNPDQASAP-----DPRPLREERARLLPRTHLQ 127  
Db 416 ETAPTTPKLPTTPEKLAAPTPEKAPTTPEELAPTTPEEPPTTPEPAPTTPEKAAAP 475  
Qy 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTPLLLELQKLPLVHATLSTPNP 180  
Db 476 NTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 531  
RESULT 5  
US-10-124-557-84  
; Sequence 84, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Bridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 84:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1022 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 84:  
US-10-124-557-84

Query Match 14.5%; Score 144; DB 13; Length 1022;  
Best Local Similarity 28.4%; Pred. No. 0.0091;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASPPRLKKEEAPLLPRHLQAEPHQHGCWTVTEPAAMT-----PGNTTTP 71  
Db 437 TTTKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 496

Qy 72 RTPEVTLRLLEQLKPLGLASTTSTNPDTQASASP-----DPRPLREEEARLLPRHLQ 127  
Db 497 ETAPTTKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAPTT 556

Qy 128 AELHQHGCWTVTEPAALT---PGNATPPTQEVTPLELLEQLKPLVHATLSTNP 180  
Db 557 NTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 612

RESULT 6  
US-10-124-557-74  
Sequence 74, Application US/10124557  
Publication No. US20020137894A1  
GENERAL INFORMATION:  
APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822

FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 74:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1038 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 74:  
US-10-124-557-74

Query Match 14.5%; Score 144; DB 13; Length 1038;  
Best Local Similarity 28.4%; Pred. No. 0.0093;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASPPRLKKEEAPLLPRHLQAEPHQHGCWTVTEPAAMT-----PGNTTTP 71  
Db 421 TTTKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 480

Qy 72 RTPEVTLRLLEQLKPLGLASTTSTNPDTQASASP-----DPRPLREEEARLLPRHLQ 127  
Db 481 ETAPTTKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAPTT 540

Qy 128 AELHQHGCWTVTEPAALT---PGNATPPTQEVTPLELLEQLKPLVHATLSTNP 180  
Db 541 NTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 596

RESULT 7  
US-10-124-557-58  
Sequence 58, Application US/10124557  
Publication No. US20020137894A1  
GENERAL INFORMATION:  
APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822

```
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58
Query Match 14.5%; Score 144; DB 13; Length 1049;
Best Local Similarity 28.4%; Pred. No. 0.0094;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
QY 23 SASPPRLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
DB 464 TTTKEPATTPEKPAATPKKPAATTPKEPATTTPKEPATTTKKPAATPAKPAATPK 523
QY 72 RTPEVTLRLLEQLKPLGLASTLTSTPNPDQASAP-----DPRPLREEEEARLLPRTHLQ 127
DB 524 ETAPTPKLTPTPEKLAATTPKEPATTPEELAPTTPEPTPTTPEEPATTPKAAAP 583
QY 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTLPLLELQKLPELVHATLSTNP 180
DB 584 NTPKEPATTPEKPAATPKPAATTPKETAATTPKGTAPTTLKEPATTTPKKAP 639
RESULT 8
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104
Query Match 14.5%; Score 144; DB 13; Length 1140;
Best Local Similarity 28.4%; Pred. No. 0.01;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
QY 23 SASPPRLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
DB 555 TTTKEPATTPEKPAATPKKPAATTPKEPATTTPKEPATTTKKPAATPAKPAATPK 614
QY 72 RTPEVTLRLLEQLKPLGLASTLTSTPNPDQASAP-----DPRPLREEEEARLLPRTHLQ 127
DB 615 ETAPTPKLTPTPEKLAATTPKEPATTPEELAPTTPEPTPTTPEEPATTPKAAAP 674
QY 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTLPLLELQKLPELVHATLSTNP 180
DB 675 NTPKEPATTPEKPAATPKPAATTPKETAATTPKGTAPTTLKEPATTTPKKAP 730
RESULT 9
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 1270 amino acids
; TYPE: amino acid
; Best Local Similarity 28.4%; Pred. No. 0.012;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
; US-10-124-557-44

Query Match          14.5%; Score 144; DB 13; Length 1270;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY      23 SASGDPRLPEEREAFLPRTHLQAEHQHCWTVTEPAAMT-----PGNTTTP 71
      : : : : : : : : : : : : : : : : : : : :
Db      421 TTTKEPAPTTPKEPAPTTPKKPAPTTTPKEPAPTTTKKPAPTAKEPAPTTK 480
      : : : : : : : : : : : : : : : : : : : :

QY      72 RTEVPTLRLELQKLPGLASTLTSTPNPDQASAP----DPRPLREEREARLLPRTHLQ 127
      : : : : : : : : : : : : : : : : : : : :
Db      481 ETAPTTPKKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEPTTTTPEEPAPTTKKAAP 540
      : : : : : : : : : : : : : : : : : : : :

QY      128 AELHQHCWTVTBPAALT---PGNATPPRTQEVPTLLLELQKLPVLVHATLSTNP 180
      : : : : : : : : : : : : : : : : : : : :
Db      541 NTPKEAPATTPKPAFTTTPKEPAPTTTPEKAPTTTGTAPTTLKEPAPTTPKKAP 596
      : : : : : : : : : : : : : : : : : : : :

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match          14.5%; Score 144; DB 13; Length 1311;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLPKEEBEEAPLLPRTHLQASPHQHGCVTVTEPAAMT-----PGNTTTP 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 TTTTKEPAPTTPKBPAPTTTPKKPAPTTPKBPAPTTTPKEPAPTTTKKPAPTA PKBPAPTTPK 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 72 RTEVPTPLRLQLPGLASTTLLSTNPDFOASASP----DPRPLREEBEARLLPRTHLQ 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 522 ETAPTTPKKLTPTTPEKLAPTTPEKAPITPEELAPITPEEPITPTTEEPAPITPKAAAP 581

QY 128 AEHLQHGCWTVTBFAALT---PGNATPPRTQEVPTPLLELQKLPELVHATLSTNP 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 582 NTPKEAPTTTPKBPAPTTTPKEPAPTTTPKEATPTTTPKGTATTLTTPKEPAPTTTPKKAP 637
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Geaner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

```

Query Match	14.5%;	Score 144;	DB 13;	Length 1313;
Best Local Similarity	28.4%;	Pred. No. 0.012;		
Matches	50;	Conservative 11;	Mismatches 97;	Indels 18; Gaps 3;
Qy	23	SASPDRLPKEEAPLLPRTHLOAEHQHCQWTVPEAAWT-----PGNTTTP 71		
Db	464	TTTKEPATTPKGPAPTTPKKPAPTTTPKEPAPTTTKKPAPTAPKEPAPTPK 523		
Qy	72	RTPEVTPRLRLQKPLGLASTLTLSNPDPTQASAP-----DPRPLREEEARLLPRTHLQ 127		
Db	524	ETAPTPKLLTPTTPEKLAPTTEPKPAPTPEELAPTTPEEPTTTEEPAPTPPKAAP 583		
Qy	128	AEHLHQHCWTVTPEAALT-----PGNATPPRQEVTPLLLELOKLPVLHATLSTNP 180		
Db	584	NTPKEPAPTPKGPAPTTTPKEPAPTTTPEKAPTTTTPKGTAPTTLKEPAPTPPKCAP 639		

RESULT 12  
US-10-124-557-50  
Sequence 50, Application US/10124557  
Publication No. US20020137894A1  
GENERAL INFORMATION:  
APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Geaner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.

ZAP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICANT NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:

```

NAME: Cseriz, Luanan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

```

Query Match 14.5%; Score 144; DB 13; Length 1314;  
Best Local Similarity 28.4%; Pred. No. 0.012;

	Matches	50; Conservative	11; Mismatches	97; Indels	18; Gaps	3;
Qy	23	SASDPRLPLKEBEAFLPRHLQLAQBPQHGGCVTVTEPAAMT-----PGNTTPP	71			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	465	TITKEPATTPKEPATTPKCPATTPKGPATTPKGPATTPYKKCAPTAKEPATTPK	524			
	:	:	:	:	:	:
	:	:	:	:	:	:
Qy	72	RTPVETPLRLELQKLPGLASTTLSTNPDTQASASP----DPRPLREEBEARLLLRTHLQ	127			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	525	ETATTPPKGLTPPTPEKLAFTTPEKPATTPBELATTPPEPTPTTPEEPATTPPKAAAP	584			
	:	:	:	:	:	:
	:	:	:	:	:	:
Qy	128	AELHQHGCWTVTEPAALT---PGNATPPRTQTVPPLLLLELQKLPELVHATLSTPNP	180			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	585	NTPKEPATTPKPEPATTPKGPATTPKGTATTPKGTATTPILKEPATTPKKCAP	640			
	:	:	:	:	:	:
	:	:	:	:	:	:

```

RESULT 13
US-10-124-557-46
; Sequence 46, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;
; NUMBER OF SEQUENCES: 143
;
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:

NAME: CSEIL, IUDANI  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1320 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
US-10-124-557-46

Query Match 14.5%; Score 144; DB 13; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 0.012;  
Matches 50; Conservative 11; Mismatches 97; Indels 18

QY 23 SASPDRPLKEEEEAPLLPRTHLQAEPHQHGCWTVTEPAAMT-----PGNTTPP 71





Qy 128 AELHQHCWVTEPAALT---PGNATPPRTQEVTPLLLELOKLPELVHATLSTNP 180  
Db 625 NTPKEPAPTTPEPAPTTPEPAPTTPEKETAFTTPKGTAPTTLKKEPAPTTPEKDPAP 680

Search completed: February 19, 2005, 00:11:41  
Job time : 134 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:50:22 ; Search time 39 Seconds  
(without alignments)  
461.347 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213  
Perfect score: 995  
Sequence: 1 LPVKPRLRGRPGSLRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	123	12.4	880	2 D89756	protein T23E7.2b [
2	122	12.3	294	2 A37232	mucin, tracheal (A
3	117.5	11.8	2649	2 T51023	hypothetical prote
4	117	11.8	653	2 E84682	hypothetical prote
5	116.5	11.7	216	2 I51920	mucin - rhesus mac
6	116	11.7	291	2 S27721	hypothetical prote
7	115.5	11.6	2187	2 T30826	nascent polypeptid
8	114.5	11.5	803	2 F59433	RhoGAP protein (im
9	114	11.5	1186	2 T19050	hypothetical prote
10	113	11.4	3020	2 A43932	mucin 2 precursor,
11	112	11.3	3149	1 Q8E8	BPLF1 protein - hu
12	111.5	11.2	548	2 E70546	hypothetical prote
13	111	11.2	846	2 T21700	hypothetical prote
14	109.5	11.0	530	2 A45690	transactivator EBN
15	109.5	11.0	794	2 C87437	penicillin-binding
16	109	11.0	524	2 S47143	A mating type prot
17	108.5	10.9	167	2 A33532	mucin SMUC-40 - hu
18	108.5	10.9	705	2 A35361	synapsin I splice
19	108.5	10.9	5262	2 T03454	ALR protein - huma
20	108	10.9	283	2 S35332	apidaecin 73 precu
21	108	10.9	377	2 A48018	mucin 7 precursor,
22	106.5	10.7	3942	2 T42730	Bassoon protein -
23	106	10.7	283	2 S13383	hydroxyproline-ric
24	106	10.7	306	2 T52340	cell wall-plasma m
25	106	10.7	915	2 T12526	hypothetical prote
26	106	10.7	1906	1 S68235	myosin-light-chain
27	105.5	10.6	798	2 T15336	hypothetical prote
28	105.5	10.6	884	2 T30479	hypothetical prote
29	105.5	10.6	1092	1 JN0635	neural cell adhesi

30 105 10.6 303 2 S28264 hydroxyproline-ric  
31 105 10.6 613 2 T00077 gag-like protein -  
32 105 10.6 699 2 C43674 US4 protein - huma  
33 105 10.6 7962 2 I38346 elastic titin - hu  
34 104.5 10.5 674 2 T05264 probable serine/th  
35 104.5 10.5 839 2 P75518 hypothetical prote  
36 104.5 10.5 2211 1 KFBOS coagulation factor  
37 104 10.5 610 2 S35049 mucin JERS7 - huma  
38 103.5 10.4 1881 2 H95076 zinc metalloprotei  
39 103.5 10.4 2688 2 T49477 alpha-A-crystallin  
40 103 10.4 369 2 G83434 translocation prot  
41 103 10.4 1396 2 A44453 translation initia  
42 103 10.4 2206 2 JC5280 voltage-dependent  
43 102.5 10.3 670 2 S22293 zinc finger protei  
44 102.5 10.3 731 2 T04455 hypothetical prote  
45 102.5 10.3 753 2 JQ0532 OP protein - Kenne

ALIGNMENTS

RESULT 1  
D89756  
protein T23E7.2b [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D89756  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D89756  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-880 <STO>  
C:Cross-references: UNIPROT:O17338; GB:chr\_X; PID:AAB71258.1; PID:G2435547; GSPDB:GN000  
C:Genetics:  
A:Gene: T23E7.2b  
A:Map position: X

Query Match 12.4%; Score 123; DB 2; Length 880;  
Best Local Similarity 26.9%; Pred. No. 0.21;  
Matches 45; Conservative 14; Mismatches 62; Indels 46; Gaps 7;  
Qy 35 EAPLLPRTHLOAEPHQHGWTVTEPAAMTPGNTTPRTPEVTPRLRLQLPGLAS--- 91  
Db ||| : : : : : ||| ||| : : : : :  
Db 392 EETPATPRSSVPATPTES---NLTPAPKTKPTKPTPR-TKPTPTKTPAVVEPEP 447  
Qy 92 -----TTLSTPNPDTOASAPDRPLREBEARLLPRTHLQALHGHGCVTVTEPA 142  
Db : : : : : ||| : : : : :  
Db 448 EPVAEEEPVAEPPEPEPEPAEPEAEPEAEPEAE-----BPE 487  
Qy 143 ALTPGNATPPRTQVTPVLLLS--LQKPELVH-----ATLSTPNP 180  
Db ||| : : : : : ||| : : : : :  
Db 488 PAADETATEPTAEAEPEAEVESEIEK-TEVVEEESAPPAARQSSPS 533

RESULT 2  
A37232  
mucin, tracheal (AMN-22) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 09-Jul-2004  
C:Accession: A37232  
R:Gerard, C.; Eddy Jr., R.L.; Shows, T.B.  
J. Clin. Invest. 86, 1921-1927, 1990  
A:Title: The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat  
A:Reference number: A37232; MUID:91072667; PMID:2254452  
A:Accession: A37232  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-294 <GER>



C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004  
C;Accession: S3615; S75613; S27721  
R;Malakhov, M.P.; Wada, H.; Los, D.A.; Sakamoto, T.; Murata, N.  
Plant Mol. Biol. 21, 913-918, 1993  
A;Title: Structure of a cyanobacterial gene encoding the 50S ribosomal protein L9.  
A;Reference number: S3614; MUID:93222488; PMID:8467083  
A;Accession: S3615  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-291 <NA2>  
A;Cross-references: UNIPROT:P42350; GB:D10716; NID:g217098; PIDN:BAA38818.1; PID:d100203  
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1992  
R;KaneKO, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.

A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S75613  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-291 <KAN>  
A;Cross-references: EMBL:D90912; GB:AB001339; NID:g1653228; PIDN:BAA16174.1; PID:d101890  
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 11.7%; Score 116; DB 2; Length 291;  
Best Local Similarity 26.8%; Pred. No. 0.2;  
Matches 49; Conservative 21; Mismatches 61; Indels 52; Gaps 9;

QY 13 PGLSLRLAEVSAPDRPLKEEAEPLLPRTHLOAPHQHGCWTVTEPAAMTPGNTVTPR 72  
DB 83 PTAIVDEMAKPSPSSPKESKSP-----KONHK----VVTFPAIVNPPTVTPAH 130  
QY 73 TPETVLRLQLKLPP---GLASTTILSTPNPDQAASDPDRPLREEEEARLLPRTLQA 128  
DB 131 PP--TPV---VEKSPEVEAAIAEPILTAP---ISFPSPDVPVLSLEP----- 172  
QY 129 ELRHQCWTVTEPAALTPGNATPRTQEVTPFLLELQ-----KLPELVHATLTSTPNPDN 182  
DB 173 -----TPPPAMV--NSTNQPEESAPIDELQLOFATPELPFLAVEAKPDSPEPM 220  
QY 183 QVT 185  
DB 221 AVS 223

RESULT 7  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N;Alternate names: alpha-NAC protein  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30826  
R;Yotkov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A;Title: Differential splicing-in of a proline-rich exon converts alphanac into a muscle  
A;Reference number: Z20889; MUID:96312450; PMID:8698236  
A;Accession: T30826  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-2187 <YOT>  
A;Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB187  
C;Gene: Naca  
A;Map position: 10  
A;Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
A;Note: differential splicing converts alphanac into a tissue-specific DNA-binding activ  
C;Keywords: alternative splicing; DNA binding; transcription factor

Query Match 11.6%; Score 115.5; DB 2; Length 2187;  
Best Local Similarity 20.5%; Pred. No. 1.9;  
Matches 56; Conservative 32; Mismatches 96; Indels 89; Gaps 8;



Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667; PMID:6087149  
A:Contents: annotation; protein coding region  
C:Superfamily: human herpesvirus 4 BPLF1 protein

Query Match 11.3%; Score 112; DB 1; Length 3149;  
Best Local Similarity 25.6%; Pred. No. 5;  
Matches 57; Conservative 21; Mismatches 79; Indels 66; Gaps 12;

QY 11 PRGSLTRLAEVASPDPR-----PLKE---EEAPLLPRTHQAE 48  
DB 395 PAPSTPRASSGAARDQTPRKGLGKDSPHKKPSGRRLPLSSDTDEDDQLPRTHVPPH 454

QY 49 -----PHQHGCTVTPEAAMTGNTPPT--PEVTPL-RLELQ---KLP 87  
DB 455 RPPSAARLPPVPIPIHQ-----SPPASPTP-HPAPVSTIAPSVTPSPRLPLQIPILP 507

QY 88 GLASTT-----LSTNPDTQASAP-----DPRPLREEEEARLLPRTHQAEHGHGCTV 138  
DB 508 QAAPSNPKIPLTTPSPSPTAAAPTTLTSPPTQQQPPQSAAPAPSPLLPQQQPTPSAA 567

QY 139 TEPAALTPGNATPRTQEV-TPLLELQKLPELVHATLSTNP 180

DB 568 PAPGELLQQPPPSAARAPSLPPQQQLPS-----ATPAP 604

RESULT 12

E70546

hypothetical protein RV0538 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C:Accession: E70546

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70546

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-548 <COL>

A:Cross-references: UNIPROT:O06404; GB:Z95558; GB:AL123456; NID:G3261781; PIDN:CAB08989.

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV0538

Query Match 11.2%; Score 111.5; DB 2; Length 548;  
Best Local Similarity 26.6%; Pred. No. 0.82;  
Matches 46; Conservative 10; Mismatches 84; Indels 33; Gaps 5;

QY 11 PRGSLTRLAEVASPDPRPLKEEEAPLLPRTHQAE-----HQHGCTVT 58  
DB 360 PRGTSFGVGCTPASAP-----EAPAVGVVAPVPFIPVPIIIPFPQGMQMTIP 412

QY 59 EPAAMTP---GNTTPRTPEVTPLRLQKLGLASTTLSTNPDTQASASDPRPLREE 115  
DB 413 TAPPTTPVTSATPT 472

QY 116 EARLLPRTHQAEHGHGCTVTPEAALTPGNATPRTQEV-TPLLELQKL 168

DB 473 APTTVAPTTPVAP-----TTVAPATATTTVAQPTQO--PQQPTQOMP 514

RESULT 13

T21700

hypothetical protein F3382.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T21700

R:Lennard, N.

submitted to the EMBL Data Library, January 1997

A:Reference number: Z19461

A:Accession: T21700

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-846 <WIL>

A:Cross-references: UNIPROT:O01699; EMBL:Z84574; PIDN:CAB06541.1; GSPDB:GN00019; CESP:F3

A:Experimental source: clone F33E2

C:Genetics:

A:Gene: CESP:F33E2.6

A:Map position: 1

A:Introns: 99/3; 213/3; 332/3; 525/3

Query Match 11.2%; Score 111; DB 2; Length 846;

Best Local Similarity 24.9%; Pred. No. 1.4;

Matches 47; Conservative 20; Mismatches 92; Indels 30; Gaps 7;

QY 6 PRLGPRGSLTRLAEVASDPRPLKEEEAPLL--PRTHQAEPHQHGCTVTPEAAM 63  
DB 571 PRTEPPR-----TEPPKTEAPRTVRPKTEAPMTVPRTPEPPMTEAPRTVPMTEPKT 623

QY 64 TPGNTTPTRT-----PEVTPLRLQKLGLASTTLSTNPDTQASASDPRPLREE 116  
DB 624 EPPRTAPRTVEVMTLPPETVPPNTEAPR-----TEVMTVP-PRTEPPKTEAPRTVPMTEPKT 678

QY 117 EARLLPRTHQAEHGHGCTVTPEAALTPGNATPRTQEV-TPLLELQKLPELVHATLS 176  
DB 679 EA---PWTEVMTGSRTEVMTPEPPKTEQRTAPPRT-----VSMTLPEETVPPKTE 729

RESULT 14

A45690

transactivator EBNA-2 - baboon lymphotropic virus

C:Species: baboon lymphotropic virus, herpesvirus papio HVP

C:Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

C:Accession: A45690

R:Ling, P.D.; Ryon, J.J.; Hayward, S.D.

J. Virol. 67, 2990-3003, 1993

A:Title: EBNA-2 of herpesvirus papio diverges significantly from the type A and type B E

ic motif.

A:Reference number: A45690; MUID:93267743; PMID:8398484

A:Molecule type: DNA

A:Residues: 1-530 <LIN>

A:Cross-references: UNIPROT:Q07701; GB:L11366; NID:G6690013; PIDN:AAA79034.1; PID:G30631

A:Note: sequence extracted from NCBI backbone (NCBIN:132361, NCBIPI:132362)

C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 11.0%; Score 109.5; DB 2; Length 530;

Best Local Similarity 22.4%; Pred. No. 1.1;

Matches 56; Conservative 25; Mismatches 92; Indels 77; Gaps 11;

QY 3 VKKPR-----LRGPRPGSLTRLAEVASDPRPLKEEEAPLLPRTHQAE 50  
DB 151 VPRPRMTLLRLPLRLIRPRPTYNPLSPNTSPSPIFP---QHTPLPRPTASPH 207

QY 51 OHGC-----WTVTEPAAMTGN--TTPRTPEVTPLRLQKLGLAST----TLSTNP 99  
DB 208 SHPQSEPKQTPKPTLPLGPPVSPPPPTPSIQTPHTPKPTIPSSGPHITL-TVDP 266

QY 100 -----DTQASASDPRPLREE-----EARLLPRTHQAEHGHGCTVTE 140  
DB 267 LSQPASTQGSNVSGPPFQHKQILITTLATSTPTTKQILPKTRSS-----ASMD 318

QY 141 PAALTPGNATPRTQEV-----TPLLLELQKL--PELVHATLST 177  
DB 319 PLPLPPLSTPPPPAPSTPSGIVDRDRTSPRLPGVWVPPVLPFHKLAGDILLTPSPDP 378

QY 178 PNPDNQVTIK 187

Search completed: February 18, 2005, 23:59:49  
Job time : 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:42:06 ; Search time 177 Seconds  
(without alignments)

541.010 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKKPLRGRPGSLRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	995	100.0	571	2 Q8NBL0	Q8nbl0 homo sapien
2	992	99.7	578	2 Q68CN3	Q68cn3 homo sapien
3	987	99.2	328	2 Q86TQ7	Q86tq7 homo sapien
4	987	99.2	571	2 Q6H9L7	Q6h9l7 homo sapien
5	969	97.4	294	2 Q86TW3	Q86tw3 homo sapien
6	933.5	93.8	658	2 Q95432	Q95432 homo sapien
7	842	84.6	518	2 Q8N501	Q8n501 homo sapien
8	321	32.3	308	2 Q86TW4	Q86tw4 homo sapien
9	144	14.5	1404	2 Q92954	Q92954 homo sapien
10	144	14.5	1404	2 Q9BX49	Q9bx49 homo sapien
11	132	13.3	837	2 Q7QFG1	Q7qfg1 anopheles g
12	124.5	12.5	5085	1 PCLO RAT	Q9jks6 rattus norv
13	123	12.4	880	2 Q17338	O17338 caenorhabdi
14	123	12.4	885	2 Q9TXR9	Q9txr9 caenorhabdi
15	123	12.4	908	2 Q8MXJ1	Q8mxj1 caenorhabdi
16	123	12.4	930	2 O17339	O17339 caenorhabdi
17	122	12.3	294	2 Q99322	Q99322 homo sapien
18	122	12.3	868	2 Q8SX72	Q8sx72 drosophila
19	122	12.3	957	1 IF2 SYNEL	Q8dk04 synchococc
20	122	12.3	1173	2 Q9VHC2	Q9vhc2 drosophila
21	121.5	12.2	488	2 Q8DGUI	Q8dgu1 synchococc
22	121	12.2	724	2 Q99LH3	Q99lh3 mus musculu
23	121	12.2	3105	2 Q8UZI9	Q8uzi9 cercopithec
24	121	12.2	3150	2 Q7PMD5	Q7pmd5 anopheles g
25	120.5	12.1	893	2 Q8WZS5	Q8wzs5 neurospora
26	119	12.0	708	2 Q7WAY9	Q7way9 bordetella
27	119	12.0	1237	2 Q7WYN2	Q7wyn2 acetivibrio
28	118.5	11.9	4969	2 Q8CF91	Q8cf91 mus musculu
29	118.5	11.9	5038	1 PCLO MOUSE	Q9qyx7 mus musculu
30	118.5	11.9	5165	2 Q8CF92	Q8cf92 mus musculu
31	117.5	11.8	2592	2 Q9P3J0	Q9p3j0 neurospora

32 117 11.8 653 2 Q9SL31 Q9sl31 arabidopsis  
33 117 11.8 660 2 Q84JK0 Q84jko arabidopsis  
34 116.5 11.7 216 2 Q28501 Q28501 macaca mula  
35 116.5 11.7 328 2 Q6CCW5 Q6ccw5 yarrowia li  
36 116 11.7 291 1 YD53 SYN3 P42350 synechoyst  
37 115.5 11.6 484 2 Q74P34 Q74p34 bacillus ce  
38 115.5 11.6 542 2 Q9SLR6 Q9slr6 triticum ae  
39 115.5 11.6 1054 2 Q9JMR9 Q9jmr9 mus musculu  
40 115.5 11.6 1371 2 Q8BV57 Q8bv57 mus musculu  
41 115.5 11.6 1509 1 GSRL HUMAN Q9nmz4 homo sapien  
42 115.5 11.6 2162 2 Q095T5 Q095t5 caenorhabdi  
43 115.5 11.6 2187 2 P70670 P70670 mus musculu  
44 115 11.6 582 2 Q82F59 Q82f59 streptomyce  
45 114.5 11.5 414 2 Q8NDG2 Q8ndg2 homo sapien

#### ALIGNMENTS

##### RESULT 1

Q8NBL0 PRELIMINARY; PRT; 571 AA.  
AC Q8NBL0;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Hypothetical protein PSEC0137.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta.  
RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
RA Saito K., Yamamoto S., Wakamatsu A., Nagai T., Nakamura Y.,  
RA Nagahari K., Sugano S., Isogai T.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK075445; BAC11626.1; --  
DR InterPro; IPR005533; AMOP.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF03782; AMOP; 1.  
DR Pfam; PF00090; TSP.1; 1.  
DR SMART; SM00723; AMOP; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
SQ SEQUENCE 571 AA; 63919 MW; 06079153C433D28C CRC64;

Query Match 100.0%; Score 995; DB 2; Length 571;

Best Local Similarity 100.0%; Pred. No. 1.5e-58;

Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPVKKPLRGRPGSLRLAEVSAPDPRLKEEEAPLLPRTHLOAEPHQHCWVTEP '60

Db 27 LPVKKPLRGRPGSLRLAEVSAPDPRLKEEEAPLLPRTHLOAEPHQHCWVTEP 86

Qy 61 AAMTPGNTTPTPEVTPLRLLEQLKPLGLASTLTSTPNPDQASAPDPRLREEEEARL 120

Db 87 AAMTPGNTTPTPEVTPLRLLEQLKPLGLASTLTSTPNPDQASAPDPRLREEEEARL 146

Qy 121 LPRTHLOAEHLQHCWVTEPAAALTPGNATPPRTQEVTPVTLLEQLKPLVHATLSTPNP 180

Db 147 LPRTHLOAEHLQHCWVTEPAAALTPGNATPPRTQEVTPVTLLEQLKPLVHATLSTPNP 206

Qy 181 DNQVTIK 187

Db 207 DNQVTIK 213

##### RESULT 2

Q68CN3 PRELIMINARY; PRT; 578 AA.

ID Q68CN3

AC Q68CN3;

DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein DKFZp686E0215 (Fragment).  
 GN Name=DKFZp686E0215;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cervix;  
 RG The German cDNA Consortium;  
 RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR749863; CAH18707.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP 1; 1.  
 DR SMART; SM00723; AMOP; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 578 AA; 64611 MW; 4BA9FA46801BD68B CRC64;  
 Query Match 99.7%; Score 992; DB 2; Length 578;  
 Best Local Similarity 99.5%; Pred. No. 2.3e-58;  
 Matches 186; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 60  
 DB 34 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 93  
 QY 61 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLASTTTLSTNPDTQASASDPRLPRREEEARL 120  
 DB 94 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLANTTLSTNPDTQASASDPRLPRREEEARL 153  
 QY 121 LPRTHLQAEHLHQGCWTVTEPAALTPGNATPPRTOEVTPLLELQKLPELVHATLSTNP 180  
 DB 154 LPRTHLQAEHLHQGCWTVTEPAALTPGNATPPRTOEVTPLLELQKLPELVHATLSTNP 213  
 QY 181 DNQVTIK 187  
 DB 214 DNQVTIK 220  
 RESULT 3  
 Q86TQ7 PRELIMINARY; PRT; 328 AA.  
 AC Q86TQ7  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Full-length cDNA 5-PRIME end of clone CS0DI033YJ09 of Placenta of Homo sapiens (human) (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX248770; CAD66577.1; -;

FT NON TER 1  
 FT NON TER 328  
 SQ SEQUENCE 328 AA; 36307 MW; 56C700854F62E89B CRC64;  
 Query Match 99.2%; Score 987; DB 2; Length 328;  
 Best Local Similarity 98.9%; Pred. No. 2.8e-58;  
 Matches 185; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 60  
 DB 31 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 90  
 QY 61 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLASTTTLSTNPDTQASASDPRLPRREEEARL 120  
 DB 91 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLANTTLSTNPDTQASASDPRLPRREEEARL 150  
 QY 121 LPRTHLQAEHLHQGCWTVTEPAALTPGNATPPRTOEVTPLLELQKLPELVHATLSTNP 180  
 DB 151 LPRTHLQAEHLHQGCWTVTEPAALTPGNATPPRTOEVTPLLELQKLPELVHATLSTNP 210  
 QY 181 DNQVTIK 187  
 DB 211 DNQVTIK 217  
 RESULT 4  
 Q6H9L7 PRELIMINARY; PRT; 571 AA.  
 ID Q6H9L7  
 AC Q6H9L7  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Thrombospondin and AMOP containing isthmin-like 1 protein, (Tail1) precursor.  
 GN Name=Tail1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RA Rossi V., Belfagna G., Rampazzo A., Baucé B., Danieli G.A.;  
 RT "TAIL1: an isthmin-like gene, containing type 1 thrombospondin-repeat and AMOP domain, mapped to ARVD1 critical region.";  
 RL Gene 335,101-108(2004).  
 DR EMBL; AJ583024; CAB47313.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP 1; 1.  
 DR SMART; SM00723; AMOP; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Signal.  
 FT CHAIN 1 26 Potential.  
 FT CHAIN 27 571 Tail1 protein.  
 SQ SEQUENCE 571 AA; 63906 MW; D83329D10F300C7C CRC64;  
 Query Match 99.2%; Score 987; DB 2; Length 571;  
 Best Local Similarity 98.9%; Pred. No. 5e-58;  
 Matches 185; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 60  
 DB 27 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEEAPLLPRTHLQAEPHQHGCVTTEP 86  
 QY 61 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLASTTTLSTNPDTQASASDPRLPRREEEARL 120  
 DB 87 AAMTPGNTTTPRTPEVTPLRLLEQLKPLGLANTTLSTNPDTQASASDPRLPRREEEARL 146  
 QY 121 LPRTHLQAEHLHQGCWTVTEPAALTPGNATPPRTOEVTPLLELQKLPELVHATLSTNP 180

Db 147 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVHATLSTNP 206  
Qy 181 DNQVTIK 187  
| | | | |  
Db 207 DNQVTIK 213  
| | | | |  
RESULT 5  
ID Q86TW3 PRELIMINARY; PRT; 294 AA.  
AC Q86TW3;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Full-length cDNA clone CSOD1014YN05 of Placenta of Homo sapiens  
DE (human) (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Li W.B., Gruber C., Jessee J., Polayes D.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BX248277; CAD62605.1; -.  
FT NON TER 1  
SQ SEQUENCE 294 AA; 31563 MW; 5127610A07112FC2 CRC64;  
Query Match 97.4%; Score 969; DB 2; Length 294;  
Best Local Similarity 98.9%; Pred. No. 3.9e-57;  
Matches 181; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LPVKKPLRGRPGSLTRLAESVSPDRPLKEEEAPLLPRTHLQAEHPHOGHCWTTVEP 60  
| | | | |  
Db 29 LPVKKPLRGRPGSLTRLAESVSPDRPLKEEEAPLLPRTHLQAEHPHOGHCWTTVEP 88  
| | | | |  
Qy 61 AAMTPGNATPRTPEVTPRLLELQKLPGLASTTLSTNPDTQASASDPDRLEEEEARL 120  
| | | | |  
Db 89 AAMTPGNATPRTPEVTPRLLELQKLPGLASTTLSTNPDTQASASDPDRLEEEEARL 148  
| | | | |  
Qy 121 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVHATLSTNP 180  
| | | | |  
Db 149 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVHATLSTNP 208  
| | | | |  
Qy 181 DNQ 183  
| | | | |  
Db 209 DNQ 211  
| | | | |  
RESULT 6  
ID Q95432 PRELIMINARY; PRT; 658 AA.  
AC Q95432;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Multimegabase Sequencing Group;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RA UW Multimegabase Sequencing Group;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF111168; AAD09622.1; -.  
DR Genew; HGNC:23176; THSD3.  
DR InterPro; IPR005533; AMOP.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF03782; AMOP; 1.  
DR Pfam; PF00090; TSP 1; 1.  
DR SMART; SM00723; AMOP; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 658 AA; 72519 MW; 7B1DF224645B300F CRC64;  
Query Match 93.8%; Score 933.5; DB 2; Length 658;  
Best Local Similarity 67.5%; Pred. No. 2.1e-54;  
Matches 185; Conservative 1; Mismatches 1; Indels 87; Gaps 1;  
Qy 1 LPVKKPLRGRPGSLTRLAESVSPDRPLKEEEAPLLPRTHLQAEHPHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVHATLSTNPDTQASASDPDRLEEEEARL 21  
| | | | |  
Db 27 LPVKKPLRGRPGSLTRLAESVSPDRPLKEEEAPLLPRTHLQAEHPHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVHATLSTNPDTQASASDPDRLEEEEARL 86  
| | | | |  
Qy 22 -----VSASPPRPLKE 33  
| | | | |  
Db 87 ERGAGAGKLPDTCRCPTGEGSTVQLIAPNAAADVHSGDKDSQTCIRVSASPPRPLKE 146  
| | | | |  
Qy 34 EEEAPLLPRTHLQAEHPHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPGLASTT 93  
| | | | |  
Db 147 EEEAPLLPRTHLQAEHPHOGHCWTTVEPAALTPGNATPRTQEVTPPLLELQKLPGLANTT 206  
| | | | |  
Qy 94 LSTNPDTQASASDPDRPLKEEEEARLLPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRT 153  
| | | | |  
Db 207 LSTNPDTQASASDPDRPLKEEEEARLLPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRT 266  
| | | | |  
Qy 154 TOEVTPLLELQKLPVLVHATLSTNPDTQVTK 187  
| | | | |  
Db 267 TOEVTPLLELQKLPVLVHATLSTNPDTQVTK 300  
| | | | |  
RESULT 7  
Q8N501 PRELIMINARY; PRT; 518 AA.  
ID Q8N501  
AC Q8N501;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE THSD3 protein (Fragment).  
GN Name=THSD3;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]

RC SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC033140; AAH33140.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP\_1; 1.  
 DR SMART; SM00723; AMOP; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 FT NON TER 1  
 SQ SEQUENCE 518 AA; 58281 MW; EB6B57C4D44E928 CRC64;

Query Match 84.6%; Score 842; DB 2; Length 518;  
 Best Local Similarity 98.1%; Pred. No. 2.1e-48;  
 Matches 157; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 28 PRPLKEEEAPLLPRTHLQAEAPHQHCWTVTEPAAMTGNTPPTPEVTEPLRLLEQLKP 87  
 DB 1 PRPLKEEEAPLLPRTHLQAEAPHQHCWTVTEPAAMTGNTPPTPEVTEPLRLLEQLKP 60  
 QY 88 GLASTTSTPNPDQASAPPRPLREEEEARLLPRTHLQAEAPHQHCWTVTEPAALTPG 147  
 DB 61 GLANTTSTPNPDQASADPRPLREEEEARLLPRTHLQAEAPHQHCWTVTEPAALTPG 120  
 QY 148 NATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNIK 187  
 DB 121 NATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNIK 160

## RESULT 8

Q86TW4 PRELIMINARY; PRT; 308 AA.  
 AC Q86TW4;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Full-length cDNA clone CS0DI024YA19 of Placenta of Homo sapiens  
 DE (human) (Fragment)  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX248280; CAD62608.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP\_1; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 FT NON TER 1  
 SQ SEQUENCE 308 AA; 34735 MW; 38F47CC67F7D44E1 CRC64;

Query Match 32.3%; Score 321; DB 2; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-14;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 AELHQCWTVTEPAALTPGNATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNIK 187  
 DB 1 AELHQCWTVTEPAALTPGNATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNIK 60

## RESULT 9

Q92954 PRELIMINARY; PRT; 1404 AA.  
 AC Q92954;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Megakaryocyte stimulating factor.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
 RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
 RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
 RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
 RT "Purification, Biochemical Characterization, and Cloning of a Novel  
 RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony  
 RT Stimulating Activity";  
 RT Blood 78:279-279(1991).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,  
 RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,  
 RA Jacobs K., Turner K.;  
 RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor";  
 RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegehoff J.,  
 RL Mosher D.F. (eds.);  
 RL BIOLOGY OF VITROTECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier  
 RL Science Publishers B.V. (1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
 RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
 RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
 RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; U70136; AAB09089.1; -;  
 DR HSSP; P04004; 10C0.  
 DR Genew; HGNC:9364; PRG4.  
 DR GO; GO:0008283; P:cell proliferation; TAS.  
 DR InterPro; IPR000585; Hemoepexin.  
 DR InterPro; IPR01212; Somatomedin\_B.  
 DR Pfam; PF00045; Hemoepexin; 2.  
 DR Pfam; PF01033; Somatomedin\_B; 2.  
 DR PRINTS; PR0022; SOMATOMEDINB.  
 DR SMART; SM00120; HX; 2.  
 DR SMART; SM00201; SO; 2.  
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN\_1.  
 DR PROSITE; PS00524; SMB\_1; 2.  
 SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 14.5%; Score 144; DB 2; Length 1404;  
 Best Local Similarity 28.4%; Pred. No. 0.19;  
 Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASPDPRPLKEEEAPLLPRTHLQAEAPHQHCWTVTEPAAMT-----PGNTTTP 71  
 DB 555 TTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 614  
 QY 72 RTPEVTPLELLEQLKLPGLASTTLPNPDQASAP-----DPRPLREEEEARLLPRTHLQ 127  
 DB 615 ETAPTTTKLTPPTPEKLAFTTPEKLAFTTPEKLAFTTPEKLAFTTPEKLAFTTPEKLA 674  
 QY 128 AELHQCWTVTEPAALT---PGNATPRTQEVTPLELLEQLKLPVLVHATLSTPNP 180

```

Db 675 NTPKEPAPTPKPEAPPTPKPEAPPTPKETAPTTPKGTAPTTLKEPAPPTPKKPAP 730
RESULT 10
Q9BX49
ID Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE BG174L6.2 (MSR: megakaryocyte stimulating factor).
GN Name=BG174L6.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR HSP; P04004; 1OC0.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS0524; SMB_1; 2.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDBE5 CRC64;

Query Match 14.5%; Score 144; DB 2; Length 1404;
Best Local Similarity 28.4%; Pred. No. 0.19; Mismatches 11; Gaps 3;
Matches 50; Conservative 11; Indels 18;

QY 23 SASDPRLKBEAEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PQNTTTP 71
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 TTTKEPAPTPKPEAPPTPKKPAPPTPKPEAPPTPKPEAPTTTKKPAPPTPK 614
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 RTPEVTPRLLEQLKPLGLASTLTSTPNPDQASAP-----DPRLREBEAEARLPRTHLQ 127
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 ETAPTTPKLPTTPEKLAPTTPKAPPTTPPELAPTTPPEAPTTPPEAPTTPKAAAP 674
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 AELHQHCWTVTEPAALT---PCGNATPRTQEVTPLELLEQLKPLVHATLTSTNP 180
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 NTPKEPAPTPKPEAPPTPKPEAPPTPKETAPTTPKGTAPTTLKEPAPPTPKKPAP 730

RESULT 11
Q7QFG1
ID Q7QFG1 PRELIMINARY; PRT; 837 AA.
AC Q7QFG1;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE AGCP13522 (Fragment).
GN Name=agCG56847; ORFName=ENSG0000000018269;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DB EMBL; AAAB01008846; EAA06308.1; -.

```

```

DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_PerA.
DR Pfam; PF01607; CBM14; 5.
DR PROSITE; PS50940; CHIT_BIND_II; 5.
FT NON_TER 1
FT NON_TER 837
SQ SEQUENCE 837 AA; 85371 MW; C506EF34EF2A726B CRC64;

Query Match 13.3%; Score 132; DB 2; Length 837;
Best Local Similarity 27.6%; Pred. No. 0.71;
Matches 55; Conservative 24; Mismatches 60; Indels 60; Gaps 12;

QY 11 PRPGSLTLAEV-SASDPRLKBEAEAPLLPRTHLQAEPHQHCWTVTEPAAM-----TP 65
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 PEPSTIPLTTPGSAATPKPT-----EPSTIPLT-----TPGSAATPKPTPESTIPLTTP 370
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 GNTTTPPTPE---VTPRLLEQLKPLGLASTLTSTPNPD-----TQASASDPRLREBE 116
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 GSATPKPTPEPTIPL-----TTPG-SATSEPTPEPTIPLTTPGSAATPEPT-----E 420
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 EARLLPRTHLQAEHLHQHCWTVTEPAALTTCGNATPPRTQE-----VTPLLLEQLKPL 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 PSTNIPLT-----TPGSTTPKPTPEPTIPLTTPGSAATSEVTPE 460
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 L-VHATLTSTPNPDNQVTK 187
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 LTTNIPLTTPGTASSIVE 479
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
PCLO RAT
ID PCLO RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JL71;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN Name=Piclo;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984; DOI=10.1016/S0896-6273(00)80883-1;
RA Fenster S.D., Chung W.U., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.",
RL Neuron 25:203-214(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP VAL-4688; 4688-VAL-NET-4689; NET-4689; 4690-VAL-VAL-4691;
RP 4692-GLN-ASN-4693 AND ALA-4694.
RX MEDLINE=21181819; PubMed=11285225; DOI=10.1093/emboj/20.7.1605;
RA Gebber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.";
RL EMBO J. 20:1605-1619(2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Prai, RIMS2 and profilin (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:

```

```

CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -- SIMILARITY: Contains 2 C2 domains.
CC -- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF138789; AAF07822.2; --
CC EMBL; AF227534; AAF63196.1; --
CC PDB; 1RH8; NMR; A=4635-4776.
CC RGD; 69406; PC10.
CC GO; GO:0045202; C:synapse; IDA.
CC GO; GO:0005509; F:calcium ion binding; IDA.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
CC GO; GO:0005522; F:profilin binding; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR008899; Znf_piccolo.
CC Pfam; PF00168; C2; 2.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF05715; zf-piccolo; 2.
CC SMART; SM00239; C2; 2.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00499; C2_DOMAIN_1; 1.
CC PROSITE; PS00004; C2_DOMAIN_2; 2.
CC PROSITE; PS0106; PDZ; 1.
CC 3D-structure; Alternative splicing; Calcium/phospholipid-binding;
KW Metal-binding; Repeat; Zinc; Zinc-finger.
FT DOMAIN 372 491
FT FT
FT ZN_FING 523 547
FT ZN_FING 1010 1033
FT DOMAIN 2351 2362
FT DOMAIN 4442 4536
FT DOMAIN 4653 4752
FT DOMAIN 4968 5059
FT VARSPIC 4876 4880
FT VARSPIC 4881 5085
FT MUTAGEN 4668 4668
FT MUTAGEN 4674 4674
FT MUTAGEN 4688 4688
FT MUTAGEN 4688 4689
FT MUTAGEN 4689 4689
FT MUTAGEN 4690 4691
FT MUTAGEN 4692 4693
FT MUTAGEN 4694 4694
FT SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
SQ
Query Match 12.5%; Score 124.5; DB 1; Length 5085;
Best Local Similarity 23.2%; Pred. No. 15;
Matches 42; Conservative 27; Mismatches 63; Indels 49; Gaps 6;
Qy 3 VKPRLGPRPGSLTRLAEVSASDPDRPLKEEEAPLLPRTHLQAEHQHGWVTEPAA 62
Db 659 VKQTLGPTAPAPQLPVAEALPEPAPPK-EPGGL-PEQ-----A 698
Qy 63 MTPGNTTPRTPETVPLRLLEQLKPLGLASTTLSTPNPDTQASASDPDRPLREEEEARLLP 122
Db 699 KAPVGDEVPKQPKWETRADIQ-----SSSTTKPDILSS-----QVQSOAQVKT 742
Qy 123 RTHLQAEHQHGWVTEPAAALTPGNATPTGRTQVTPLLLEQLKPLVHATLSTPNPDN 182
Db 743 ASPLKTD-----SAXPSQSFPTGKTTPLDKAMPASDASKIISQPGPGS 789
Qy 183 Q 183
Db 790 E 790

```

## RESULT 13

017338

ID 017338 PRELIMINARY; PRT; 880 AA.

AC 017338;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein T23E7.2.

GN ORFNames=T23E7.2, T23E7.2;

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;

RG WormBase Consortium;

RT "Genome sequence of the nematode C. elegans: a platform for

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Latreil P., Stellyes L., Elliot G., Wilson R.;

RT "The sequence of C. elegans cosmid T23E7.2";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Wilton R.;

RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RG WormBase Consortium;

RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF026205; AAB71258.1; --

DR PIR; D89756; D89756.

DR WormBase; WBGene00020732; T23E7.2.

DR WormPep; T23E7.2b; CE14066.

DR Hypothetical protein.

KW SEQUENCE 880 AA; 95398 MW; 97A8A101E8FBA1C1 CRC64;

SQ
Query Match 12.4%; Score 123; DB 2; Length 880;
Best Local Similarity 26.9%; Pred. No. 3;
Matches 45; Conservative 14; Mismatches 62; Indels 46; Gaps 7;







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 00:06:56 ; Search time 2780 Seconds  
(without alignments)  
3259.398 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKPRLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10030225/runat\_18022005\_100354\_2341/app.query.fasta\_1.327  
-DB=GenEmbl -QPMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225\_QCGN\_1\_1\_5600 @runat\_18022005\_100354\_2341 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb ba:\*  
2: gb htg:\*  
3: gb in:\*  
4: gb om:\*  
5: gb ov:\*  
6: gb pat:\*  
7: gb ph:\*  
8: gb pl:\*  
9: gb pr:\*  
10: gb ro:\*  
11: gb sts:\*  
12: gb sy:\*  
13: gb un:\*  
14: gb vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
1	995	100.0	2981 6 AX136231 Sequence
2	995	100.0	2981 6 BD093285 Different
3	995	100.0	2981 6 BD123568 Secretary
4	995	100.0	2981 9 AK075445 Homo sapi

5	987	99.2	2883	9	AJ583024	Homo sapi
6	964.5	96.9	3052	9	AK056709	Homo sapi
7	938.5	94.3	1698	6	CQ720015	Sequence
8	857.5	86.2	750	6	AX136531	Sequence
9	857.5	86.2	750	6	BD123771	Secretary
10	842	84.6	2783	9	BC033140	Homo sapi
c 11	841.5	84.6	231464	9	AF111168	Homo sapi
c 12	658.5	66.2	497	6	CQ049341	Sequence
c 13	658.5	66.2	497	6	CQ064365	Sequence
c 14	658.5	66.2	497	6	CQ091307	Sequence
c 15	658.5	66.2	497	6	CQ130141	Sequence
c 16	658.5	66.2	497	6	CQ168762	Sequence
c 17	658.5	66.2	497	6	CQ197880	Sequence
c 18	658.5	66.2	497	6	CQ213333	Sequence
c 19	658.5	66.2	497	6	CQ251912	Sequence
c 20	658.5	66.2	497	6	CQ289061	Sequence
c 21	658.5	66.2	497	6	CQ326064	Sequence
c 22	430	43.2	243	6	CQ054407	Sequence
c 23	430	43.2	243	6	CQ073647	Sequence
c 24	430	43.2	243	6	CQ104532	Sequence
c 25	430	43.2	243	6	CQ143244	Sequence
c 26	430	43.2	243	6	CQ178734	Sequence
c 27	430	43.2	243	6	CQ203077	Sequence
c 28	430	43.2	243	6	CQ226434	Sequence
c 29	430	43.2	243	6	CQ264575	Sequence
c 30	430	43.2	243	6	CQ301666	Sequence
c 31	430	43.2	243	6	CQ338879	Sequence
c 32	234.5	23.6	172662	10	AC120540	Mus muscu
c 33	154.5	15.5	189436	2	AL691432	Homo sapi
c 34	153.5	15.4	146664	9	AC010327	Homo sapi
c 35	150	15.1	123255	9	AC006003	Homo sapi
c 36	150	15.1	184132	2	AC079590	Homo sapi
c 37	150	15.1	200070	9	AC008746	Homo sapi
c 38	149.5	15.0	6694	6	AX252013	Sequence
c 39	149.5	15.0	6694	6	AX348784	Sequence
c 40	149	15.0	132211	9	AC091196	Homo sapi
c 41	149	15.0	170880	2	AC145870	Sequence
c 42	148	14.9	183930	2	AC141015	Rattus no
c 43	147	14.8	6456	6	AX345908	Sequence
c 44	146	14.7	132634	2	AC123294	Rattus no
c 45	146	14.7	171849	9	AC010616	Homo sapi

ALIGNMENTS

RESULT 1  
AX136231  
LOCUS AX136231  
DEFINITION Sequence 153 from Patent EP1067182.  
ACCESSION AX136231  
VERSION AX136231.1 GI:14272639  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and  
Hayaishi, K.  
TITLE Secretary protein or membrane protein  
JOURNAL Patent: EP 1067182-A 153 10-JAN-2001;  
Helix Research Institute (JP)  
FEATURES  
Location/Qualifiers  
source  
1..2981  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
58..1773  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39758.1"  
/db\_xref="GI:14272640"  
/translation="MRALRDRAGLLLCVLLAALLAALGPGVKKPRLRGRPGSLTR"

LAEVASDPDRPLKEEBEAPLLPRTHLOAEPHOHCWTVTEPAAMTNGTTPRTTPEV  
 TPLRLKQLPGLASTTSLTPNDPTQASAPDPRLPREEBEARLLPRTHLOAELHOGH  
 CWTVEPAALTPGNATPRTQEVTPPLLELQKLPVLVATLSLPNDQVTVIKVREDP  
 QAEVSDLLAEPPSPDDTSLSPALWPFIMWDYKGEKDRAPGEGKEBEDEDYP  
 SEDTEGEDQEDKEDEEALWFGTTDNDQGLAFDGVFVXDSVSYDYEPQKELRN  
 WSPGSGNCSTQKQRTPCGTCGTATRTCDLPSCPGTGDKDTGLSPSEWKLARL  
 ATMDHDDVDSCEKWLKDFLTKYLSQMLRDLSPCPAPKPLEAMDSPLVQDEHOG  
 RSPWRDASGPREDLDIYQPTARECLRSMLSGESSTIAAOHCVCDESRLLTRGKAG  
 MPNLISDFSPKLFHKFDTPFWILCKGDSRLHVLFPNNGRACDTPLEEEVLAQLQ  
 EAKY"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,67e-40 Length: 2981  
 Score: 995.00 Matches: 187  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x AX136231 (1-2981)

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 136 CTCCCGGTGAAGAGCCGCGCTCCGCGGACACCGGCTGGGAGCCTCACGAGGCTCGCA 195  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 196 GAGGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGGAGGACCACTGCTC 255  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
 Db 256 CCCAGAACCCACCTCGAGCGAGGACCAACCAATGATGCTGACCTGTCACTGAGCCA 315  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
 Db 316 GCAGCATGACCCAGGCAACACCACTCCAGGACCCAGAGGTTACTCCGTTGGG 375  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 376 CTGGAGCTGCAGAGCTCCCGGATGTCAGCAACCTTGGTACCCCTAACCTTGAT 435  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluAlaArgLeu 120  
 Db 436 ACCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGACGACGACTG 495  
 Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 Db 496 CTCCCGAGAACCCACCTCGAGCGAGGCTACACCAACATGGATGTTGACCTGTCACTGAG 555  
 Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProArgThrGlnGluValThrProLeu 160  
 Db 556 CCAGCAGGCTTACCCCGAGGAAATGCCAGCTCCCGAGACCCAGAGGTTACTCCCTTG 615  
 Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 616 CTGCTGGAGCTGCAGAGCTGCAGATTTGGTCCACGCAACCTTTCAGTACCCCTAACCT 675  
 Qy 181 AspAsnGlnValThrIleLys 187  
 Db 676 GATAACCAAGGTACCATCAAG 696

## RESULT 2

BD093285  
 LOCUS 2981 bp DNA linear PAT 27-AUG-2002  
 DEFINITION Differentiation growth factor.  
 ACCESSION BD093285  
 VERSION BD093285.1 GI:22638873  
 KEYWORDS WO 0104312-A/1.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2981)  
 AUTHORS Oca,T., Isoqai,T., Nishikawa,T., Kawai,Y., Yoshida,K. and Masuho,Y.  
 TITLE Differentiation growth factor  
 JOURNAL Patent: WO 0104312-A 1 18-JAN-2001;  
 HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,  
 YURI KAWAI, KENJI YOSHIDA, YASUHIKO MASUHO  
 COMMENT OS Homo sapiens (human)  
 PN WO 0104312-A/1  
 PD 18-JAN-2001  
 PF 06-JUL-2000 WO 2000JP004514  
 PR 08-JUL-1999 JP 99P 194179, 18-OCT-1999 US 60/159586 PI  
 TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, KENJI PI  
 YOSHIDA,  
 PI YASUHIKO MASUHO  
 PC C12N15/16, C12N15/12, C12N15/85, C12N5/10, C12P21/02, C07K14/575,  
 C07K14/72  
 PC C12Q1/68, C12Q1/02, A61K67/027// (C12P21/02, C12R1:91) CC  
 FH Key Location/Qualifiers  
 FT CDS (58)..(1770).

## FEATURES

Source  
 1. .2981  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,67e-40 Length: 2981  
 Score: 995.00 Matches: 187  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BD093285 (1-2981)

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 136 CTCCCGGTGAAGAGCCGCGCTCCGCGGACACCGGCTGGGAGCCTCACGAGGCTCGCA 195  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 196 GAGGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGGAGGAGGACCACTGCTC 255  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
 Db 256 CCCAGAACCCACCTCGAGCGAGGACCAACCAATGATGCTGACCTGTCACTGAGCCA 315  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
 Db 316 GCAGCATGACCCAGGCAACACCACTCCAGGACCCAGAGGTTACTCCGTTGGG 375  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 376 CTGGAGCTGCAGAGCTCCCGGATGTCAGCAACCTTGGTACCCCTAACCTTGAT 435  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluAlaArgLeu 120  
 Db 436 ACCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGACGACGACTG 495  
 Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 Db 496 CTCCCGAGAACCCACCTCGAGCGAGGCTACACCAACATGGATGTTGACCTGTCACTGAG 555  
 Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProArgThrGlnGluValThrProLeu 160  
 Db 556 CCAGCAGGCTTACCCCGAGGAAATGCCAGCTCCCGAGACCCAGAGGTTACTCCCTTG 615  
 Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 616 CTGCTGGAGCTGCAGAGCTGCAGATTTGGTCCACGCAACCTTTCAGTACCCCTAACCT 675  
 Qy 181 AspAsnGlnValThrIleLys 187

```
Db 676 GATAACCAAGGTCACCATCAAG 696
|||||
BD123568
LOCUS BD123568 2981 bp DNA linear PAT 18-SEP-2002
DEFINITION Secretory protein or membrane protein.
ACCESSION BD123568
VERSION BD123568.1 GI:23218513
KEYWORDS JP 2002017376-A/77.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2981)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: JP 2002017376-A 77 22-JAN-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002017376-A/77
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253173
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,TOMOYASU
PI SUGIYAMA,
PI KOJI HAYASHI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,C12N15/00,C12N5/00 CC
Secretory protein or membrane protein
FH' Key Location/Qualifiers
FT CDS (58)..(1770).
FEATURES
source Location/Qualifiers
1..2981
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 1.67e-40 Length: 2981
Pred. No.: 995.00 Matches: 187
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6
US-10-030-225-2_COPY_27_213 (1-187) x BD123568 (1-2981)
Qy 1 LeuProVallyLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
|||||
Db 136 CTCCTCCGCGTGAAGACGCGCGCTCCGGGACACCGCTGGAGCCTCAGAGGCTCGCA 195
|||||
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
|||||
Db 196 GAGTCTTCAGCTCCACAGATCCTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 255
|||||
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
|||||
Db 256 CCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 315
|||||
Qy 61 AlaAlaValThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
|||||
Db 316 GCAGCCATGACCCAGGAGCAACACCCCTCCAGGACCCAGAGGTTACTCCGTCGGG 375
|||||
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
|||||
Db 376 CTGAGCTGCAGAGCTGCCGGGATTGCCAGGACCAACCTTGAGTACCCCTAACCTGAT 435
|||||
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
|||||
```

```
436 ACCAGGCTTCAGCCTCCCCAGATCCTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGCTG 495
121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
|||||
496 CTCCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
|||||
141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
|||||
556 CCAGCAGCCTGACCCCGAGGAAATGCCACGCTCCAGGACCCAGGAGGTTACTCCCTTG 615
|||||
161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
|||||
616 CTGCTGAGGCTGCAGAAAGCTGCCAGAAATGGTCACGCAACCTTGAGTACCCCTAACCC 675
|||||
181 AspAsnGlnValThrLeuLys 187
|||||
676 GATAACCAAGGTCACCATCAAG 696
|||||
RESULT 4
AK075445
LOCUS Homo sapiens cDNA PSEC0137 fis, clone PLACE1005047. linear PRI 03-SEP-2002
DEFINITION AK075445
ACCESSION AK075445
VERSION AK075445.1 GI:22761538
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Kawai-Hio,Y., Hayashi,K.,
Ishii,S., Saito,K., Yamamoto,J., Wakamatsu,A., Nagai,T.,
Nakamura,Y., Nagahara,K., Sugano,S. and Isogai,T.
TITLE HRI human cDNA sequencing project
JOURNAL Unpublished
JOURNAL 2 (bases 1 to 2981)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES
source Location/Qualifiers
1..2981
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1005047"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"
58..1773
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC11626.1"
/db_xref="GI:22761539"
/translation="MRALDRAGLLLCVLLAALLLEAALGLPVKKPLRGRPRGSLTR
LAIVSASDPDRPLKEEELPLPRTHLQAEHPDPRPLREBEARLLPRTHLQAEHLQHG
TPRLLEQLQPLGLASTLTSTNPDTQASDPDRPLREBEARLLPRTHLQAEHLQHG
CWTVTPAALTPGNATPPRTQEVTPLLLEQLQPLGLWGTGKEEDKDRAPGEKEEEDYP
QAEVSDLLAEPSNPPQDTLSWLPALWFLWGDYKGEEDKDRAPGEKEEEDYP
SEDIIEGQDQEDBEELQALWNTGTTDNDQWGLAPGDFVFDVSVSYDEYQKEWSP
WSPSGNCTSCQKQRTPCGYGCTATRTCDLPSCFGTCDTGLTPSPSEWKLARN
ATDMHQDVDSCKWLNKSDFLIKYLSOMLRDLPCPCAYPLFAMDSVYSLQDLHQG
RSFRWRDAGPRERLDIYQPTARFLKRLSMLSGSSSTAQHCYCCYEDSRLLTRGKAG
MPNLISDTSPFKLHFEDTTPWILCKGDSRLHVLFPNNGRACCTDNPLEEYLAQLQ
EAKY"
```

Alignment Scores:  
 Pred. No.: 1,67e-40 Length: 2981  
 Score: 995.00 Matches: 187  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x AK075445 (1-2981)

Qy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 136 CTCGCCGTGAAGAGCGCGCTCCGCGGACACACGCGCTCGGAGCCTCACGAGGCTGCA 195  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 196 GAGTCTCAGGCTCCCGAGATCTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 255  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 Db 256 CCCAGAACCCCTCGCAGCGAGGCCACCAACATGGATGCTGGACTGTCACTGAGCCA 315  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 Db 316 GCAGCATGACCCAGGCAACACACCCCTCCAGGACCCAGAGGTTTACTCCGTTGCGG 375  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 376 CTGGAGCTGCAAGAGCTCCGCGGATGCGCCAGCAACCTTGAGTACCCCTTAACCCCTGAT 435  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 Db 436 ACCCAGGCTTCAGCTCCCGAGATCTAGGCTCTGAGGGAAGAGGAGGACGACGACTG 495  
 Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTTPThrValThrGlu 140  
 Db 496 CTCGCCAGAACCCAGCTCGAGGAGCTACACCAACATGGATGTTGACTGTCACTGAG 555  
 Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 Db 556 CCAGCAGCCCTGACCCCGAGGAATGCCAGCGCTCCAGGACCCAGAGGAGTTTACTCCCTTG 615  
 Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 616 CTGCTGGAGCTGCAGAACTGCCAGATGTCACCGCAACTTGAGTACCCCTAACCCCT 675  
 Qy 181 AspAsnGlnValThrIleLys 187  
 Db 676 GATAACCAAGGTGACCATCAAG 696

RESULT 5  
 AJ583024  
 LOCUS Homo sapiens mRNA for Tail1 protein. linear PRI 18-JUN-2004  
 DEFINITION  
 ACCESSION AJ583024  
 VERSION AJ583024.1 GI:48958174  
 KEYWORDS Tail1 gene; Tail1 protein; thrombospondin and AMOP containing ischmin-like 1 protein.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 Rossi,V., Belfagna,G., Rampazzo,A., Bauce,B. and Danieli,G.A.  
 AUTHORS Rossi,V.  
 TITLE Tail1: an ischmin-like gene, containing type 1 thrombospondin-repeat and AMOP domain, mapped to ARVD1 critical region  
 JOURNAL Gene 335, 101-108 (2004)  
 PUBMED 15194193  
 REFERENCE 2 (bases 1 to 2883)  
 AUTHORS Rossi,V.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-SEP-2003) Rossi V., Department of Biology, University

of Padua, viale G. Colombo, 3 - Padova, 35131, ITALY

# FEATURES

## source

1..2883  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /map="14q24.3"  
 /tissue\_type="heart"  
 /dev\_stage="adult"  
 1..2883  
 /gene="TAIL1"  
 36..1751  
 /gene="TAIL1"  
 /functions="putative role in development"

/codon\_start=1  
 /product="thrombospondin and AMOP containing ischmin-like 1 protein, (Tail1)"

1 protein, (Tail1)"  
 /protein\_id="CAE47313.1"

/db\_xref="GI:48958175"

/translation="MRALRDAGLLLCVLLAALLAALGLPVKKPRLRGPRPSLFR

LAESASPDRPLKEEEELPRLTHIQABPHQGCWTVTEPAAMTPGNATPPRTPV

TPURLEQLPLGANTTSTPNPTQASDPRLREEREARLLPRLTHQAEUHQH

CWTEPALTPGNATPPRTQEVTPLLLEQLKLPVLVHATLSTPNQNTIKVDEP

QAEVSIIDLAEPNPPQDITSLPALWLSLWGYKGEKDRAPGEKEEDEDYP

SEDIEGEDQEKDEEEQALWNGTTDNDQGLAPGWDVKDSVYDVEPOKWS

WSPCNCSTCKQRTPCGYGTATETRTCDLPSCPGTDEKDTLGLPSEWKLARN

ATDHQDVDSCKWLNKSDFLIKYLSQMLRDLPCFAYPLEAMDSPVSLQEHQ

RSRPRDASGPRRLDIYPTAPFLRSLMSGESSTLAQHCCYDEDSRLTLRGKAG

MPNLISDFSKLHFKFDTTPWILKGDWSRLHAVLPNNNGRACTDNPLEEYLAQLQ

EAKY"

36..113  
 /gene="TAIL1"

114..1748  
 /gene="TAIL1"

/product="Tail1 protein"

2865..2870  
 /gene="TAIL1"

## sig\_peptide

## mat\_peptide

## polyA\_signal

## ORIGIN

Alignment Scores:  
 Pred. No.: 4.02e-40 Length: 2883  
 Score: 987.00 Matches: 185  
 Percent Similarity: 99.47% Conservative: 1  
 Best Local Similarity: 98.93% Mismatches: 1  
 Query Match: 99.20% Indels: 0  
 DB: 9 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x AJ583024 (1-2883)

Qy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 114 CTCGCCGTGAAGAGCGCGCTCCGCGGACACACGCGCTCGGAGCCTCACGAGGCTGCA 173  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 174 GAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 233  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 Db 234 CCCAGAACCCAGCTCGAGGAGGAGGCCACCAACATGGATGCTGACTGTCACTGAGCCA 293  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 Db 294 GCAGCCATGATCCCGAGCAACGCCCTCCAGGACCCAGAGGTTTACTCCGTTGCGG 353  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 354 CTGGAGCTGCAGAGCTGCCGGATTGGCCACACACACTTGAGTACCCCTAACCCCTGAT 413  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 Db 414 ACCCAGGCTTCAGCCTCCCGAGATCTTAGGCTCTGAGGGAAGAGGAGGACGACGACTG 473



```

Db      642 CCAGGGAATGCCACGCCCTCCACGAGACCAGAGAGGTACTCCCTTGCTGCTCGAGCTGCAG 701
Qy      166 LysLeuProGluLeuValHisAlaThrLeuSerThrProAsnProAspAsnGlnValThr 185
Db      702 AAGCTGCCAGAAATGGTCCAGCAACCTTGAGTACCCCTAACCTTGATACCAAGGTGACC 761
Qy      186 IleLys 187
Db      762 ATCAAG 767

RESULT 7
CQ720015
LOCUS      CQ720015          1698 bp      DNA          linear      PAT 03-FEB-2004
DEFINITION Sequence 5949 from Patent WO02068579.
ACCESSION  CQ720015
VERSION    CQ720015.1 GI:42280872
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS   Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE     Kite, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL   Patent: WO 02068579-A 5949 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES   source
            Location/Qualifiers
            1..1698
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      6.2e-38      Length:      1698
Score:          938.50      Matches:    186
Percent Similarity: 68.25%      Conservative: 1
Best Local Similarity: 67.88%      Mismatches: 0
Query Match:    94.32%      Indels:     87
DB:              6          Gaps:         1

US-10-030-225-2_COPY_27_213 (1-187) x CQ720015 (1-1698)

Qy      1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db      79 CTCCTCCGTGAAGAGCCGCGCTCGCGGACACCGGCTTGGAGCTTACGAGGCTCGCA 138
Qy      21 Glu----- 21
Db      139 GAGGTGAGTGGCGGGCCGCGGCTCGGAGCGGCTGCGTCCACCTCCCGAGCCC 198
Qy      21 ----- 21
Db      199 GCCGGCTCCAGCCGTGGGGAGCGGACCGGGACACACAGCAGCGACCCCGCATG 258
Qy      21 ----- 21
Db      259 GAGAGAGGGCGGGCGGGCGGGAAGCTCCCGACACCGGGCGTGGCCAGTCACTGAG 318
Qy      21 ----- 21
Db      319 GGGTCTACAGTCAGCTTATTGCCCTTGAATGCAGGAGCGTTTCATTCCTCATGGGAC 378
Qy      22 -----ValSerAlaSerProAspProArgProLeuLysGlu 33
Db      379 AAAGACAGCCAGACTTGTATTAGGGTCTCAGCCCTCCCGAGATCTTAGGCTCTGAAGGAA 438
Qy      34 GluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHisGly 53
Db      439 GAGGAGGAGGACCACTGCTCTCCCGAGAACCCACCTTGCAGGAGAGCCACACCAACATGGA 498

```

```

Qy      54 CysTrpThrValThrGluProAlaAlaMetThrProGlyAsnThrThrProProArgThr 73
Db      499 TGCTGGACTGTCACTGAGCCAGCAGCCATGACCCCGAGCAACACACCCCTCCCGAGACC 558
Qy      74 ProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThr 93
Db      559 CCAGAGGTTACTCGTTGCGGCTGGAGCTGCAGAAAGCTGCGGGATTGGCCACACACACC 618
Qy      94 LeuSerThrProAsnProAspThrGlnAlaSerAlaSerProAspProArgProLeuArg 113
Db      619 TTGAGTACCCCTAACCTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTCGAGG 678
Qy      114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db      679 GAAGAGGAGGAGGAGGACGACTGCTCCCGCAGAACCCACCTGCGAGGCAGAGCTACACCAACAT 738
Qy      134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db      739 GGATGTTGGACTGTCACTGAGCCAGCAGCCCTGACCCCGAGGAATGCGACGCTCCCGAGG 798
Qy      154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db      799 ACCGAGGAGGTTACTCCTTGTCTGCTGAGCTGCGAAGCTGCCAGAATTGGTCCACGCA 858
Qy      174 ThrLeuSerThrProAsnProAspAsnGlnValThrIleLys 187
Db      859 ACCTTGAGTACCCCTAACCTGATACCCAGGTGACCATCAAG 900

RESULT 8
AX136531
LOCUS      AX136531          750 bp      DNA          linear      PAT 30-MAY-2001
DEFINITION Sequence 453 from Patent EP1067182.
ACCESSION  AX136531
VERSION    AX136531.1 GI:14272935
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS   Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.
TITLE     Secretory protein or membrane protein
JOURNAL   Patent: EP 1067182-A 453 10-JAN-2001;
Helix Research Institute (JP)
FEATURES   Location/Qualifiers
            source
            1..750
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      2.96e-34      Length:      750
Score:          857.50      Matches:    175
Percent Similarity: 93.65%      Conservative: 2
Best Local Similarity: 92.59%      Mismatches: 6
Query Match:    86.18%      Indels:     6
DB:              6          Gaps:         1

US-10-030-225-2_COPY_27_213 (1-187) x AX136531 (1-750)

Qy      1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db      136 CTCCTCCGTGAAGAGCCGCGCTCCCGGACACCGGCTTGGAGCTTACGAGGCTCGCA 195
Qy      21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db      196 GAGGTCTCAGCCCTCCCGAGATCTTAGGCTCTGAAGGAGGAGGAGGAGGACCACTGCTC 255
Qy      41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db      256 CCCAGAACCCACCTGCGAGGAGGAGCCACACCAATGGATGCTGGACTGTCACTGAGCCA 315

```



```

Qy 61 AlalaMetThrProGlyAsnThrThrProProArgThrThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGACCCAGGCAACACACCCCTCCAGGACCCAGAGGTACTCCGTTGGG 375

Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAla-SerThrThrLeuSerThrProAsnProAs 100
Db 376 CTGAGCTGCGAGAGCTGCCGGGATGGCCAAAGCACAACTTGAGTACCCCTAACCCCTGA 435

Qy 100 pThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLe 120
Db 436 TACCAGGCTTCAGCTCCCGAGATCTTANGCCTCTGAGGGAAGAGGAGGAGGACGACT 495

Qy 120 uLeuProArgThrHisLeuGlnAla-GluLeuHisGlnHisGlyCysTrpThrValThrG 140
Db 496 GCTCCCGAGAACCCCTGCAGGCAAGAGCTACACCAACATGGATGTGGACTGTCACTG 555

Qy 140 luProAlaAlaLeuThrProGlyAsnAlaThrPro-ProArgThrGlnGluValThrPro 159
Db 556 AGCCAGCAGAGCTTCAGCCAGGGAATGCCAGCTCCCAAGACCCCAAGANGTTACTCCC 615

Qy 160 LeuLeuLeuGluLeuGlnLys---LeuProGluLeuValHisAla-ThrLeuSer-ThrP 178
Db 616 TTGCTGCTGGAGCTGCAGAAAGCTGCCAGAAATGGTCCACGCAAACTTGAGTTANCC 675

RESULT 9
BD123771
LOCUS
DEFINITION
SECRETORY protein or membrane protein.
ACCESSION
BD123771
VERSION
JP 2002017376-A/280.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 '(bases 1 to 750)
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
SECRETORY protein or membrane protein
Patent: JP 2002017376-A 280 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017376-A/280
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253173
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,TOMOYASU
PI SUGIYAMA,
PI KOJI HAYASHI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC
SECRETORY protein or membrane protein
FH Key Location/Qualifiers
FTI source 1..750
FT Location/Qualifiers
1..750
/organism="Homo sapiens (human)"
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2.96e-34 Length: 750
Score: 857.50 Matches: 175
Percent Similarity: 93.65% Conservative: 2
Best Local Similarity: 92.59% Mismatches: 6

```

```

Query Match: 86.18% Indels: 6
DB: 6 Gaps: 1
US-10-030-225-2_COPY_27_213 (1-187) x BD123771 (1-750)

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 136 CTCCCCGTGAAGAGCCGCGCTCCGCGACCCAGCGCTGGAGCCTCACAGAGCTCGCA 195

Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 196 GAGGTCTCAGCTCCCGAGATCTTAGGCTCTGAGGAAGAGGAGGAGGACCACTGCTC 255

Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 256 CCCAGAACCCACCTGCAGGCGAGCCACCAACATGGATGCTGACTGTCACTGAGCCA 315

Qy 61 AlalaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCAGAGGTACTCCGTTGGG 375

Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAla-SerThrThrLeuSerThrProAsnProAs 100
Db 376 CTGGAGCTGCAGAGCTGCCGGGATGGCCAAAGCACAACTTGAGTACCCCTAACCCCTGA 435

Qy 100 pThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLe 120
Db 436 TACCAGGCTTCAGCTCCCGAGATCTTANGCCTCTGAGGGAAGAGGAGGAGGACGACT 495

Qy 120 uLeuProArgThrHisLeuGlnAla-GluLeuHisGlnHisGlyCysTrpThrValThrG 140
Db 496 GCTCCCGAGAACCCCTGCAGGCAAGAGCTACACCAACATGGATGTGGACTGTCACTG 555

Qy 140 luProAlaAlaLeuThrProGlyAsnAlaThrPro-ProArgThrGlnGluValThrPro 159
Db 556 AGCCAGCAGGCTTCAGCCAGGGAATGCCAGCTCCCAAGGACCCCAAGANGTTACTCCC 615

Qy 160 LeuLeuLeuGluLeuGlnLys---LeuProGluLeuValHisAla-ThrLeuSer-ThrP 178
Db 616 TTGCTGCTGGAGCTGCAGAAAGCTGCCAGAAATGGTCCACGCAAACTTGAGTTANCC 675

Qy 178 roAsnProAspAsnGln 183
Db 676 CTAACCTGTAAACCAA 692

RESULT 10
BD123771
LOCUS
DEFINITION
Homo sapiens thymospondin, type I, domain containing 3, mRNA
(GDNA clone IMAGE:3844836), partial cds.
ACCESSION
BD123771
VERSION
BD123771.1
GI:21619879
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 '(bases 1 to 2783)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

```





```

repeat_region      /rpt_family="L1MD1"
200. .494
/rpt_family="AluSx"
repeat_region      500. .1283
/rpt_family="AluSg"
repeat_region      1284. .1581
/rpt_family="AluSg"
repeat_region      1582. .1774
/rpt_family="AluJb"
repeat_region      1775. .2058
/rpt_family="AluJb"
repeat_region      2059. .2934
/rpt_family="L191_Send"
repeat_region      complement(2937. .3255)
/rpt_family="MER4A"
repeat_region      3323. .3687
/rpt_family="L1"
repeat_region      3688. .3977
/rpt_family="AluSg"
repeat_region      4009. .4314
/rpt_family="AluJo"
repeat_region      4321. .4934
/rpt_family="L1"
repeat_region      4940. .5236
/rpt_family="AluSp"
repeat_region      5238. .5349
/rpt_family="L1MC1"
repeat_region      5350. .5650
/rpt_family="AluSq"
repeat_region      5888. .5924
/rpt_family="TAAAA)n"
repeat_region      5869. .5940
/rpt_family="CAAAA)n"
repeat_region      6086. .6271
/rpt_family="AluJb"
repeat_region      complement(6500. .6588)
/rpt_family="5S"
unsure             complement(6715. .6815)
note="low quality data"
repeat_region      complement(6914. .7458)
/rpt_family="LINE2"
repeat_region      7523. .7707
/rpt_family="L1MB6"
repeat_region      7708. .8007
/rpt_family="AluSx"
repeat_region      8233. .8305
/rpt_family="TA)n"
repeat_region      9186. .9230
/rpt_family="CA)n"
CDS               complement(join(9248. .9470,10060. .10240,16529. .16722,
25330. .25484))
/note="intron-exon boundaries defined in relation to an
EST contigs that includes AA397580, A1138314, AA399545,
AA68184, all from testis. The closest match in BLASTX is
to E. coli pyruvate-5-carboxylate reductase."
/codon_start=1
/product="unknown"
/protein_id="A0D09625.1"
/db_xref="GI:4186186"
/translation="MDMLQDLESIQFYGVPEEDRIWLYQGRSGLMIEACAHATFF
CKLYNLRASLNKNQSRHLSIGLSNATPEEPKVGIIQGGHGLKQLAGTLQIGTIP
AESLRITRRPETIGELQKGIKGFYHNADLVSWADVFLLCPLSPQPNICVETITSL
EKASIVSPVAALPLPLKLLNHTNLRQYQYDEDSVSWGANKGVIALQDPITL
QATCPYPAGGSGTYFVHSPENYFPVGPLLSY"
complement(9631. .9708)
repeat_region      /rpt_family="LINE2"
repeat_region      complement(10461. .10762)
/rpt_family="AluSx"
repeat_region      complement(10804. .11077)
/rpt_family="AluJb"
repeat_region      complement(11156. .11306)
/rpt_family="MIR"
repeat_region      complement(11493. .11615)
/rpt_family="FLAM_C"
11800. .12161
/rpt_family="THEIC"
complement(12406. .12695)
/rpt_family="AluSx"
12746. .12920
/rpt_family="BC200"
12921. .13217
/rpt_family="AluSx"
13236. .13751
/rpt_family="L1MB1"
13783. .14084
/rpt_family="AluSx"
14085. .14403
/rpt_family="L1MB2"
14405. .14717
/rpt_family="AluSx"
14795. .14857
/rpt_family="L1MC2"
complement(14884. .14905)
/rpt_family="AT rich"
complement(15106. .15393)
/rpt_family="AluSx"
complement(15606. .15818)
/rpt_family="L1MB2"
complement(15835. .16131)
/rpt_family="AluSc"
complement(16775. .17073)
/rpt_family="AluY"
17491. .17777
/rpt_family="AluJo"
17778. .17804
/rpt_family="CA)n"
complement(17815. .17882)
/rpt_family="L1MB2"
17883. .18175
/rpt_family="AluSx"
18177. .18251
/rpt_family="TAGA)n"
complement(18244. .18306)
/rpt_family="L1MB1"
complement(18310. .18609)
/rpt_family="AluSx"
complement(18624. .18771)
/rpt_family="L1MB3"
complement(18765. .18847)
/rpt_family="L1"
18858. .19501
/rpt_family="SVA"
complement(19503. .20417)
/rpt_family="L1"
complement(20104. .20625)
/rpt_family="L143_Send"
complement(20724. .21029)
/rpt_family="AluJb"
complement(21030. .21122)
/rpt_family="TAAA)n"
complement(21125. .21423)
/rpt_family="AluJo"
complement(21486. .21662)
/rpt_family="L191_Send"
complement(21664. .21966)
/rpt_family="AluSq"
complement(22042. .22342)
/rpt_family="AluY"
complement(22350. .22937)
/rpt_family="L143_Send"
complement(22975. .23090)
/rpt_family="FLAM_C"
complement(23164. .23287)
/rpt_family="L143_Send"
complement(26429. .26704)
/rpt_family="L1MB5"

```





CQ130141/c	CQ130141	497 bp	DNA	linear	PAT 21-JAN-2004
LOCUS	Sequence 163 from Patent WO0157276.				
DEFINITION	CQ130141				
ACCESSION	CQ130141				
VERSION	CQ130141.1	GI:41087497			
KEYWORDS	· Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.				
TITLE	Human genome-derived single exon nucleic acid probes useful for analysis of gene expression in human bone marrow				
JOURNAL	Patent: WO 0157276-A 163 09-AUG-2001;				
FEATURES	Agemica, Inc. (US)				
source	Location/Qualifiers				
	1..497				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
	/note="MAP TO AF111168.2-EXPRESSED IN BONE MARROW, SIGNAL = 0.92"				

## ORIGIN

Alignment Scores:			
Pred. No.:	1.27e-24	Length:	497
Score:	658.50	Matches:	126
Percent Similarity:	76.97%	Conservative:	1
Best Local Similarity:	76.36%	Mismatches:	1
Query Match:	66.18%	Indels:	37
DB:	6	Gaps:	1
US-10-030-225-2_COPY_27_213 (1-187) x CQ130141 (1-497)			
Qy	32	LysGluGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln	51
Db	496	AAGGAAGAGGAGGAGGACCACTGCTCCCGACAGACCCACCTGCAGGCAGAGCCACCA	437
Qy	52	HisGlyCysThrThrValThrGluProAlaAlaMetThrProGlyAsnThrThrProPro	71
Db	436	CATGATCTCGACTGTCCTAGCAGCAGCGCATGACCCCGAGCAAGCCACCCCTCCC	377
Qy	72	ArgThrProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSer	91
Db	376	AGGACCCAGAGGTTACTCCGTTGGCGCTGGAGCTGCAGAAAGCTGCGGGATTGCCAAC	317
Qy	92	ThrThrLeuSerThrProAsnProAspThr	101
Db	316	ACAACCTTGAGTACCCTTAACCTCATACCAGGTGAGAGCTACAGAGGGCCAGCAGCT	257
Qy	101	-----	101
Db	256	CCTGGCTCTGCCTGCCAGCCCTTTCTGGTCCCTGTGTGGTAAGCCCTAACCTGTCAAC	197
Qy	102	-----GlnAlaSerAlaSerProAspProArgProLeuArgGlu	114
Db	196	TCACCTTCCCTCTCTCTTCTCCTCAGGCTTCAGCCTCCCCAGATCTTAGGCCTCTGAGGAA	137
Qy	115	GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly	134
Db	136	GAGGAGGAGGACGACGATGCTCCCGACAGACCCACCTGCGAGCGAGAGCTACCAACATGA	77
Qy	135	CysThrThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArgThr	154
Db	76	TGTTGACATGTCATGAGCCAGCAGCCCTGACCCCGAGGAAATGCCAGGCTCCAGGACC	17
Qy	155	GlnGluValThrPro	159
Db	16	CAGGAGGTTACTCCC	2

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 00:04:01 ; Search time 423 Seconds  
(without alignments)

2617.001 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKPLRGPRGSLRLLA.....PELVHATLSTPNPDNQVTK 187

Scoring table:

BLOSUM62  
-Q/cgn2 1/USPTO pool p/US10030225/runat 18022005 100354 2333/app query.fasta\_1.327  
-DB=US Geneseq 16Dec04 -DPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPCU=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10030225 @CGN 1 1 708 @runat 18022005 100354 2333 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2 1/USPTO pool p/US10030225/runat 18022005 100354 2333/app query.fasta\_1.327  
-DB=US Geneseq 16Dec04 -DPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPCU=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10030225 @CGN 1 1 708 @runat 18022005 100354 2333 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:\*

1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002as:.\*  
7: Geneseqn2002bs:.\*  
8: Geneseqn2003as:.\*  
9: Geneseqn2003bs:.\*  
10: Geneseqn2003cs:.\*  
11: Geneseqn2003ds:.\*  
12: Geneseqn2004as:.\*  
13: Geneseqn2004bs:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	995	100.0	2981	4	Asf293148 Human pro
2	995	100.0	2981	5	Asf93820 Human CDN
3	857.5	86.2	750	5	Asf94019 Primer sp
C 4	658.5	66.2	497	4	Aai10232 Probe #16
C 5	658.5	66.2	497	4	Abas1868 Human foe

C	6	558.5	66.2	497	4	AAI31480
C	7	658.5 <td>66.2</td> <td>497</td> <td>4</td> <td>ABA21690</td>	66.2	497	4	ABA21690
C	8	658.5 <td>66.2</td> <td>497</td> <td>4</td> <td>Aak25606 Human bon</td>	66.2	497	4	Aak25606 Human bon
C	9	658.5 <td>66.2</td> <td>497</td> <td>4</td> <td>Aak00167 Human bra</td>	66.2	497	4	Aak00167 Human bra
C	10	658.5 <td>66.2</td> <td>497</td> <td>4</td> <td>Abas25182 Human liv</td>	66.2	497	4	Abas25182 Human liv
C	11	658.5 <td>66.2</td> <td>497</td> <td>5</td> <td>Aai00170 Probe #16</td>	66.2	497	5	Aai00170 Probe #16
C	12	658.5 <td>66.2</td> <td>497</td> <td>6</td> <td>Abas00175 Human gen</td>	66.2	497	6	Abas00175 Human gen
C	13	430	43.2	243	4	Aai19514 Probe #94
C	14	430	43.2	243	4	Abas4531 Human foe
C	15	430	43.2	243	4	Aai44705 Probe #13
C	16	430	43.2	243	4	Abas46659 Human bre
C	17	430	43.2	243	4	Abas31664 Probe #10
C	18	430	43.2	243	4	Aak38709 Human bon
C	19	430	43.2	243	4	Aak12982 Human bra
C	20	430	43.2	243	4	Abas38283 Human liv
C	21	430	43.2	243	5	Aai05236 Probe #52
C	22	430	43.2	243	6	Abas12780 Human gen
C	23	149.5	15.0	6694	6	Abel70352 Chemical
C	24	149.5	15.0	6694	6	Aas61313 Human gen
C	25	147	14.8	6456	6	Abel33006 Human imm
C	26	145.5	14.6	74787	12	Adq74672 Streptomy
C	27	144	14.5	3813	10	Adl13808 Osteoarth
C	28	144	14.5	3936	10	Adl13807 Osteoarth
C	29	144	14.5	4092	10	Adl13806 Osteoarth
C	30	144	14.5	4140	12	Adk67916 Human ext
C	31	144	14.5	4215	10	Adl13804 Osteoarth
C	32	144	14.5	4215	10	Adl13805 Osteoarth
C	33	144	14.5	4289	12	Adk67917 Human ext
C	34	144	14.5	4575	8	Abx63303 Human cdn
C	35	144	14.5	4847	10	Adk65818 Angiogene
C	36	144	14.5	5008	2	Aaq27223 MSF precu
C	37	144	14.5	5012	10	Adl13810 Osteoarth
C	38	144	14.5	5041	8	Abz34656 Coding se
C	39	144	14.5	5041	8	Abz34732 Coding se
C	40	144	14.5	5041	8	Abz34731 Coding se
C	41	144	14.5	5041	8	ACA64826 Human meg
C	42	144	14.5	5041	10	Adl12655 Human cdn
C	43	144	14.5	5041	10	Adl13803 Osteoarth
C	44	144	14.5	5041	11	Adl131655 Human cdn
C	45	144	14.5	5041	12	Adm98015 Human meg

#### ALIGNMENTS

RESULT 1

AAF29348

ID AAF29348 standard; cDNA; 2981 BP.

XX

AC AAF29348;

XX

DT 20-APR-2001 (first entry)

XX

XX Human proliferation differentiation factor cDNA sequence.

DE

XX Human; proliferation differentiation factor; haematopoietic function; ss.

XX Homo sapiens.

XX

XX W0200104312-A1.

XX

PD 18-JAN-2001.

XX

PF 06-JUL-2000; 2000WO-JP004514.

XX

PR 08-JUL-1999; 99JP-00194179.

XX

PR 18-OCT-1999; 99US-0159586P.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

XX

DR WPI; 2001-138354/14.

DR

P-PSDB; AAB49765.

XX	Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cDNA library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.
PT	Human cDNA encoding a membrane or secretory protein clone PSEC0137.
DE	Human, secretory protein; membrane protein; vaccine; gene therapy; rheumatoid arthritis; diabetes; ss.
XX	Homo sapiens.
OS	EPI067182-A2.
PN	10-JAN-2001.
XX	07-JUL-2000; 2000EP-00114090.
PR	08-JUL-1999; 99JP-00194179.
PR	11-JAN-2000; 2000JP-00118775.
PR	02-MAY-2000; 2000JP-00183766.
XX	(HELI-) HELIX RES INST.
OT	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
WFI	WFI; 2001-093989/11.
DR	P-PSDB; AAB88393.
XX	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development.
PT	Claim 1; SEQ ID NO 153; 609pp + Sequence Listing; English.
XX	This invention relates to nucleic acid sequences AAF93744 - AAF93916 which encode human secretory or membrane proteins represented by AAB88317 - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and AAF62235 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes
XX	Sequence 2981 BP; 609 A; 925 C; 887 G; 560 T; 0 U; 0 Other;
SQ	Alignment Scores:
PA	Pred. No.: 2,15e-50 Length: 2981
XX	Score: 995.00 Matches: 187
XX	Percent Similarity: 100.00% Conservative: 0
XX	Best Local Similarity: 100.00% Mismatches: 0
XX	Query Match: 100.00% Indels: 0
DB:	Gaps: 5
US-10-030-225-2_COPY_27_213 (1-187) x AAF93820 (1-2981)	
QY	1 LeuProValIysLysProArgLeuArgGlyProArgProCysSerLeuThrArgLeuAla 20
Db	136 CTCCTCCGTGAAGAAGCGCGCTCCGGACCAACGCCCTCCAGGCCTCACAGGCTCGCA 195
QY	21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db	196 GAGTCTCAGCTCCCCAGATCTTAGGCTCTCAAGAAAGAGGAGGACCACCTTGCTC 255
QY	41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db	256 CCCAAGAACCCACTGCAGGAGAGGCCACCAACATGGATGCTGGACTGTCTCAGGCCA 315
QY	61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db	316 GCAGCCATGACCCCAGGAAACACACCCCTCCAGAACCCAGAGGTACTTCGGTGGG 375
QY	81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db	376 CTGGAGCTGCAGAAGCTGCGGGATTGCCAGACACCACTTGAGTACCCCTAACCCCTGAT 435
QY	101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db	436 ACCCAGGCTTACGCTCCCAAGATCCCTAGGCTCTGAGGAGAGGAGGAGGACCATG 495
QY	121 LeuProArgThrHisLeuGlnAlaGluHisGlnHisGlyCysTrpThrValThrGlu 140
Db	496 CTCCTCCAGAACCCACTGCAGGAGAGCTACCAACATGGATGTTGGACTGTCACTGAG 555
QY	141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db	556 CCAGCAGCCCTGACCCAGGGAATGCCAGCCTCCAGAGCCAGGAGGTACTTCCTCTG 615
QY	161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db	616 CTGCTGGAGCTGCAGAACCTGCCAGAAATGGTCCACCACTTGAGTACCCCTAACCT 675
QY	181 AspAsnGlnValThrIleLys 187
Db	676 GATAACCAAGGTACCATCAAG 696
RESULT 2	
AAF93820	
ID	AAF93820 standard; cDNA; 2981 BP.
XX	AAF93820;
XX	

```
Db 196 GAGGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAGAGGAGGACCACTGCTC 255
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 256 CCCAGAACCCACCTGTCAGGAGAGGACCAACCAATGATGCTGGACTGTCACTGAGGCA 315
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGATGACCCAGGCAACCAACCCCTCCAGGACCCAGAGGTTACTCCGTTGGGG 375
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 376 CTGAGCTGCGAAGCTCCGGGATGCGGAGCAACCTTGAGTACCCCTAACCTGAT 435
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 436 ACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGACGACTG 495
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 496 CTCGCCAGAACCCACCTGTCAGGAGAGCTACCAACATGGATGTTGACTGTCACTGAG 555
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 556 CCAGAGCCCTGACCCCGAGGAATGCCAGCTCCAGGACCCAGAGGTTACTCCCTTG 615
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 616 CTGCTGGAGCTGCAGAGCTGCCAGATTGGTCCACGCAACCTTGAGTACCCCTAACCT 675
Qy 181 AspAsnGlnValThrIleLys 187
Db 676 GATAACCAAGTGACCATCAAG 696

RESULT 3
AAF94019 standard; DNA; 750 BP.
XX AAF94019;
AC AAF94019;
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 453.
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes; PCR primer; ss.
XX Synthetic.
XX
FN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-00114090.
XX
PR 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI; 2001-093989/11.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.
XX
XX Claim 4; SEQ ID NO 453; 609pp + Sequence Listing; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
```

```
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes
XX
```

Sequence 750 BP; 170 A; 249 C; 215 G; 109 T; 0 U; 7 Other;

#### Alignment Scores:

Pred. NO.:	8.67e-43	Length:	750
Score:	857.50	Matches:	175
Percent Similarity:	93.65%	Conservative:	2
Best Local Similarity:	92.59%	Mismatches:	6
Query Match:	86.18%	Indels:	6
DB:	5	Gaps:	1

US-10-030-225-2\_COPY\_27\_213 (1-187) x AAF94019 (1-750)

```
Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 136 CTCCTCCGTGAAGAGCGCGGCTCCGGAGCACCGCTGGAGCCTCAGAGGCTCGCA 195
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 196 GAGGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAGAGGAGGACCACTGCTC 255
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 256 CCCAGAACCCACCTGTCAGGAGGAGGACCAACCAATGGATGCTGACTGTCACTGAGCA 315
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGATGACCCAGGCAACCAACCCCTCCAGGACCCAGAGGTTACTCCGTTGGGG 375
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAla-SerThrThrLeuSerThrProAsnProAs 100
Db 376 CTGAGCTGCGAAGCTCCGGGATGCGGAGCAACCTTGAGTACCCCTAACCTGA 435
Qy 100 pThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 436 TACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGACGACT 495
Qy 120 uLeuProArgThrHisLeuGlnAla-GluLeuHisGlnHisGlyCysTrpThrValThrG 140
Db 496 GCTCCCGAGAACCCACCTGCGAGGCAAGCTACCAACATGGATGTTGACTGTCACTG 555
Qy 140 luProAlaAlaLeuThrProGlyAsnAlaThrPro-ProArgThrGlnGluValThrPro 159
Db 556 AGCCAGAGCCCTGACCCCGAGGGAATGCCAGCTCCAGGACCCAGAGGTTACTTCCC 615
Qy 160 LeuLeuLeuGluLeuGlnLys---LeuProGluLeuValHisAla-ThrLeuSer-ThrP 178
Db 616 TTGCTGCTGGAGCTGCAAGAAAGCTGCCAGAAATTGGTCCACGCAACCTTGAAGTANCC 675
Qy 178 roAsnProAspAsnGln 183
Db 676 CTAACCTGATACCAAC 692
```

RESULT 4  
AAAI10232/c  
AAI10232 standard; DNA; 497 BP.  
XX  
XX  
AC  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Probe #165 for gene expression analysis in human cervical cell sample.  
XX  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
cervical cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157278-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US000670.  
PP  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-488901/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human cervical epithelial cells.  
FT  
XX Claim 25; SEQ ID NO 165; 487bp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging of  
CC diseases of the cervix, notably cervical cancer. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 497 BP; 81 A; 128 C; 177 G; 111 T; 0 U; 0 Other;  
SQ  
  
Alignment Scores:  
Pred. No.: 4,11e-31 Length: 497  
Score: 658.50 Matches: 126  
Percent Similarity: 76.97% Conservative: 1  
Best Local Similarity: 76.36% Mismatches: 1  
Query Match: 66.18% Indels: 37  
DB: 4 Gaps: 1  
  
US-10-030-225-2\_COPY\_27\_213 (1-187) x AAI10232 (1-497)  
  
Qy 32 LysGluGluGluAlaProLeuLeuProGThrHisLeuGlnAlaGluProHisGln 51  
Db 496 AAGGAGAGAGAGAGGACCACTGCTCCAGACCCACCTGCAGGAGAGCCACCA 437  
Qy 52 HisGlyCysrThrValThrGluProAlaMetThrProGlyAsnThrProPro 71  
Db 436 CATGGATGCTGGACTGTCTACTGAGCAGCAGCATGACCCAGGCAAGCCACCTCC 377  
Qy 72 ArgThrProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAla 91



```
XX SQ Sequence 497 BP; 81 A; 128 C; 177 G; 111 T; 0 U; 0 Other;
Alignment Scores:
Score: 4.11e-31 Length: 497
Pred. No.: 658.50 Matches: 126
Percent Similarity: 76.97% Conserv: 1
Best Local Similarity: 76.36% Mismatches: 1
Query Match: 66.18% Indels: 37
DB: 4 Gaps: 1

US-10-030-225-2_COPY_27_213 (1-187) x ABA51868 (1-497)
Oy 32 LysGluGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db 496 AAGGAAGAGGAGGAGGACCACTGCTCCCAAGAACCCACCTGCAGGAGAGCCACCA 437
Oy 52 HisGlyCysTrpThrValThrGluProAlaAlaMetThrProGlyAenThrProPro 71
Db 436 CATGGATGCTGGACTGTCCTACTGAGCCAGCAGCCATGACCCAGGCAACGCCCTCC 377
Oy 72 ArgThrProGluValThrProLeuArgLeuGluLeuLysLeuProGlyLeuAlaSer 91
Db 376 AGGACCCAGAGGTTACTCCGTTGGGCTGGAGTGCAGAAAGCTGCCGGGATTGCCAAC 317
Oy 92 ThrThrLeuSerThrProAsnProAspThr----- 101
Db 316 ACAACCTTGAGTACCCCTTAACCTGATACCCAGGTGAGAGCTACAGAAAGGCCAGCAGCT 257
Oy 101 ----- 101
Db 256 CCTGGCTCTGCTGGCCAGCCCTTTTCTGGGTCCTGTTGGTAAAGCCCTAACCTGTGCACC 197
Oy 102 -----GlnAlaSerAlaSerProAspProArgProLeuArgGlu 114
Db 196 TCACCTTCCCTCTCTCTCTCCTCAGCTTCCCAAGAACCCACCTGCAGGAGTCTAGG 137
Oy 115 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly 134
Db 136 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 77
Oy 135 CysTrpThrValThrGluProAlaAlaLeuThrProGlyAenAlaThrProProArgThr 154
Db 76 TGTGGACTGTCACTGAGCCAGCAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAGC 17
Oy 155 GlnGluValThrPro 159
Db 16 CAGGAGGTTACTCCC 2

RESULT 6
AAI31480/c
ID AAI31480 standard; DNA; 497 BP.
XX AC AAI31480;
XX XX
DT 17-OCT-2001 (first entry)
XX XX
DE Probe #166 used to measure gene expression in human placenta sample.
XX XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX OS Homo sapiens.
XX PN W0200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
```



ABS25182/C

RESULT 10  
ABS25182/9





```
Pred. No.: 4.11e-31 Length: 497
Score: 658.50 Matches: 126
Percent Similarity: 76.97% Conservative: 1
Best Local Similarity: 76.36% Mismatches: 1
Query Match: 66.18% Indels: 37
DB: 6 Gaps: 1

US-10-030-225-2_COPY_27_213 (1-187) x ABS00175 (1-497)

Qy 32 LysGluGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db 496 AAGGAAGAGGAGGAGCCACCTGCTCCCGAGAACCCCTGCGAGGAGGCCACACCA 437
Qy 52 HisGlyCysTTPThrValThrGluProAlaAlaMetThrProGlyAsnThrThrProPro 71
Db 436 CATGATGCTGGAGTGTCTAGTGGAGCAGCAGCCATGACCCAGGCAACGCCACCTCC 377
Qy 72 ArgThrProGluValThrProLeuArgLeuGluGlnLysLeuProGlyLeuAlaSer 91
Db 376 AGGACCCAGAGGTTACTCCGTTGGCTGGAGCTGCAGAAAGCTGCCGGGATTGGCCA 317
Qy 92 ThrThrLeuSerThrProAsnProAspThr----- 101
Db 316 ACAACCTTGAGTACCCCTAACCTGATACCCAGGTGAGAGCTACAGAAAGGCCAGCT 257
Qy 101 ----- 101
Db 256 CTGGCTCTGCTGGCCAGCCTTTCTGGGTCCCTGTTGGTAAGCCCTAACCTGTAC 197
Qy 102 -----GlnAlaSerAlaSerProAspProArgProLeuArgGlu 114
Db 196 TCACCTTCCCTCTCTCTCTCAGGCTTCAGCTCCCGAGATCCTTAGGCTCTGAGGGA 137
Qy 115 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly 134
Db 136 GAGGAGGAGGACGACTGCTCCCGAGAACCCACCTGCGAGGAGAGCTACACCAACATGA 77
Qy 135 CysTTPThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArgThr 154
Db 76 TGTGGACTGTCTGACGACGAGCCCTGACCCAGGGATGTCACCCCTCCAGGACC 17
Qy 155 GlnGluValThrPro 159
Db 16 CAGGAGGTACTCCC 2

RESULT 13
AA119514/c
ID AA119514 standard; DNA; 243 BP.
AC AA119514;
XX
DT 12-OCT-2001 (first entry)
DE
DE Probe #9447 for gene expression analysis in human cervical cell sample.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.
Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human cervical epithelial cells.
Claim 25; SEQ ID NO 9447; 487pp; English.
The present invention relates to human single exon nucleic acid probes
(SENPs). The present sequence is one such probe. The SENPs are derived
from human HeLa cells. The SENPs can be used to produce a single exon
microarray, which can be used for measuring human gene expression in a
sample derived from human cervical epithelial cells. By measuring gene
expression, the probes are therefore useful in grading and/or staging of
diseases of the cervix, notably cervical cancer. Note: The sequence data
for this patent did not form part of the printed specification, but was
obtained in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 243 BP; 34 A; 61 C; 89 G; 59 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.28e-18 Length: 243
Score: 430.00 Matches: 79
Percent Similarity: 98.77% Conservative: 1
Best Local Similarity: 97.53% Mismatches: 1
Query Match: 43.22% Indels: 0
DB: 4 Gaps: 0

US-10-030-225-2_COPY_27_213 (1-187) x AA119514 (1-243)
Qy 22 ValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeuPro 41
Db 243 GTCTCAGCTCCCGAGATCCTTAGGCTCTGAAGGAGGAGGAGGAGGAGGAGGAGG 184
Qy 42 ArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 61
Db 183 AGAACCCACCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 124
Qy 62 AlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeu 81
Db 123 GCCATGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 64
Qy 82 GluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrThrProAsnProAspThr 101
Db 63 GAGCTGCAGAAAGCTGCCGGGATTGGCCAAACACCAACCTTGAGTACCCCTAACCTGATACC 4
Qy 102 Gln 102
Db 3 CAG 1

RESULT 14
ABA64531/c
ID ABA64531 standard; DNA; 243 BP.
XX
AC ABA64531;
XX
DT 01-FEB-2002 (first entry)
DE
DE Human foetal liver single exon nucleic acid probe #12836.
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
```

[illegible]

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 01:18:16 ; Search time 151 Seconds  
(without alignments)  
2026.382 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKPRLRGRPGSLRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool\_p/US10030225/runat\_18022005\_100355\_2367/app.query.fasta\_1.327  
-DB=issued\_patents\_NA -QMT=fastcap -SUFFIX=nni -MINMATCH=0.1 -LOOPCD=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225 @CIGN 1.1 105 @runat\_18022005\_100355\_2367 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	15.1	767677	4	US-09-949-016-12147
2	150	15.1	767677	4	Sequence 12147, A
3	144	14.5	2824	3	US-09-949-016-17361
4	144	14.5	3066	3	Sequence 13, Appl
5	144	14.5	3117	3	Sequence 83, Appl
6	144	14.5	3148	3	Sequence 73, Appl
7	144	14.5	3420	3	Sequence 57, Appl
8	144	14.5	3813	3	Sequence 103, Appl
9	144	14.5	3936	3	Sequence 43, Appl
10	144	14.5	3942	3	Sequence 41, Appl
11	144	14.5	3945	3	Sequence 141, Appl
12	144	14.5	3963	3	Sequence 49, Appl
					Sequence 45, Appl

13	144	14.5	3963	3	US-07-757-022B-59	Sequence 59, Appl
14	144	14.5	4065	3	US-07-757-022B-47	Sequence 47, Appl
15	144	14.5	4086	3	US-07-757-022B-39	Sequence 39, Appl
16	144	14.5	4092	3	US-07-757-022B-51	Sequence 51, Appl
17	144	14.5	4215	3	US-07-757-022B-61	Sequence 61, Appl
18	144	14.5	4847	4	US-10-164-595-57	Sequence 57, Appl
19	144	14.5	5008	3	US-07-757-022B-1	Sequence 1, Appl
20	144	14.5	5017	4	US-09-949-016-4956	Sequence 4956, Ap
21	144	14.5	5041	4	US-09-023-655-981	Sequence 981, App
22	144	14.5	5041	4	US-09-298-970A-2	Sequence 2, Appl
23	144	14.5	21707	4	US-09-949-016-16698	Sequence 16698, A
24	138.5	13.9	38239	4	US-09-949-016-12348	Sequence 12348, A
C 24	138.5	13.9	38252	4	US-09-949-016-13570	Sequence 13570, A
C 25	138.5	13.9	38252	4	US-09-949-016-13570	Sequence 13570, A
26	136.5	13.7	59853	4	US-09-949-016-13618	Sequence 13618, A
27	136.5	13.7	59853	4	US-09-949-016-13619	Sequence 13619, A
28	136.5	13.7	59853	4	US-09-949-016-13620	Sequence 13620, A
29	136.5	13.7	59853	4	US-09-949-016-13621	Sequence 13621, A
30	136.5	13.7	59853	4	US-09-949-016-13622	Sequence 13622, A
31	136.5	13.7	59853	4	US-09-949-016-13623	Sequence 13623, A
32	136.5	13.7	59853	4	US-09-949-016-13624	Sequence 13624, A
33	136.5	13.7	59853	4	US-09-949-016-13625	Sequence 13625, A
34	135	13.6	18722	4	US-09-949-016-15892	Sequence 15892, A
35	133	13.4	15695	4	US-09-949-016-15644	Sequence 15644, A
36	132	13.3	2130	4	US-09-909-962A-7	Sequence 7, Appl
C 37	131.5	13.2	80161	3	US-09-036-987A-1	Sequence 1, Appl
C 38	131.5	13.2	80161	3	US-09-370-700-1	Sequence 1, Appl
C 39	131.5	13.2	80161	4	US-09-603-207-1	Sequence 1, Appl
C 40	130	13.1	14462	4	US-09-902-540-1090	Sequence 1090, Ap
C 41	130	13.1	14462	4	US-09-902-540-9597	Sequence 9597, Ap
C 42	127.5	12.8	601	4	US-09-949-016-198377	Sequence 198377, A
C 43	127.5	12.8	92344	4	US-09-949-016-16802	Sequence 16802, A
C 44	127	12.8	27630	4	US-09-949-016-16362	Sequence 16362, A
C 45	126.5	12.7	1131	6	5168049-2	Patent No. 5168049

ALIGNMENTS

RESULT 1

US-09-949-016-12147  
; Sequence 12147, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12147  
; LENGTH: 767677  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(767677)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12147

Alignment Scores:  
Pred. No.: 4.93  
Score: 150.00  
Percent Similarity: 42.00%  
Best Local Similarity: 32.00%  
Query Match: 15.08%  
DB: 4  
Length: 767677  
Matches: 64  
Conservative: 20  
Mismatch: 63  
Indels: 53  
Gaps: 11

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-09-949-016-12147 (1-767677)

Qy 7 ArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSerPro 26  
Db 604963 CAGTTACACAGGACCGAGCGGCTTCTTATAAATGCACAGCCACGTCAGTTACACCA 605022

Qy 27 AspProArgProLeuLysGluGluGluAla-----Pro 38  
Db 605023 GGACCGACCGAGCTTCTCATATAATGCACAGCCACGTCAGTTACACAGCAGCTACCCG 605082

Qy 39 LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHis-----GlyCysTrpThr 56  
Db 605083 GCTCTCTTACAAACCCACAGCGCGTGTCAGTTACACAGGACCGCGGCTTCTTATAAA 605142

Qy 57 ValThrGluProAlaAlaMetThrProGlyAsnThrThrProArg-----72  
Db 605143 AAGCAGAGCGCGCAACAGTTACACAGGA-----CGACCGAGCTTCTTATAAA 605193

Qy 73 ---ThrProGluVal-----ThrPro 78  
Db 605194 TGCACAGCCCGAGTCAGTTACACAGCCGCGGCTTCTTATAAATGCACAGCCG 605253

Qy 99 ProAspThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAla 118  
Db 605254 CGTCAGTTACAC--CAGCACCTACCGGC-----TCTCTTACAAACCCACAGCC 605349

Qy 119 ArgLeu-----LeuProArgThrHisLeuGlnAlaGluLeu 130  
Db 605350 CGCGTCAGTTACACAGCAGCTACCGCGCTCTCTTATAAATGCACAGCCGCGTCAGTTA 605409

Qy 131 HisGlnHis-----GlyCysTrpThrValThrGluProAlaAlaLeuThr-----145  
Db 605410 CACACAGCCGCGGCTTCTTATAAATGCACAGCCGCGTCAGTTACACAGCACCT 605469

Qy 146 -ProGlyAsnAlaThrProProArgThrGlnGluValThrProLeuLeuLeuLeu 164  
Db 605470 ACCCGGCTCTCTTACAAACCCACAGCCGCGTGTCAGTTACACAGCATTGACCCGCGCTC 605527

RESULT 2  
US-09-949-016-17361  
; Sequence 17361, Application US/09949016  
; Patent No. 681239  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17361  
; LENGTH: 767677  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)\_(767677)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17361

Alignment Scores:

Pred. No.: 4.93 Length: 767677  
Score: 150.00 Matches: 64  
Percent Similarity: 42.00% Conservative: 20  
Best Local Similarity: 32.00% Mismatches: 63  
Query Match: 15.08% Indels: 53  
DB: 4 Gaps: 11

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-09-949-016-17361 (1-767677)

Qy 7 ArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSerPro 26  
Db 604963 CAGTTACACAGGACCGAGCGGCTTCTTATAAATGCACAGCCACGTCAGTTACACCA 605022

Qy 27 AspProArgProLeuLysGluGluGluAla-----Pro 38  
Db 605023 GGACCGACCGAGCTTCTCATATAATGCACAGCCACGTCAGTTACACAGCAGCTACCCG 605082

Qy 39 LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHis-----GlyCysTrpThr 56  
Db 605083 GCTCTCTTACAAACCCACAGCGCGTGTCAGTTACACAGGACCGCGGCTTCTTATAAA 605142

Qy 57 ValThrGluProAlaAlaMetThrProGlyAsnThrThrProArg-----72  
Db 605143 AAGCAGAGCGCGCAACAGTTACACAGGA-----CGACCGAGCTTCTTATAAA 605193

Qy 73 ---ThrProGluVal-----ThrPro 78  
Db 605194 TGCACAGCCCGAGTCAGTTACACAGCCGCGGCTTCTTATAAATGCACAGCCG 605253

Qy 99 ProAspThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAla 118  
Db 605254 CGTCAGTTACAC--CAGCACCTACCGGC-----TCTCTTACAAACCCACAGCC 605298

Qy 119 ArgLeu-----LeuProArgThrHisLeuGlnAlaGluLeu 130  
Db 605350 CGCGTCAGTTACACAGCAGCTACCGCGCTCTCTTATAAATGCACAGCCGCGTCAGTTA 605409

Qy 131 HisGlnHis-----GlyCysTrpThrValThrGluProAlaAlaLeuThr-----145  
Db 605410 CACACAGCCGCGGCTTCTTATAAATGCACAGCCGCGTCAGTTACACAGCACCT 605469

Qy 146 -ProGlyAsnAlaThrProProArgThrGlnGluValThrProLeuLeuLeuLeu 164  
Db 605470 ACCCGGCTCTCTTACAAACCCACAGCCGCGTGTCAGTTACACAGCATTGACCCGCGCTC 605527

RESULT 3  
US-07-757-022B-13  
; Sequence 13, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2824 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2823  
US-07-757-022B-13

Alignment Scores:  
Pred. No.: 0.00919 Length: 2824  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-13 (1-2824)

Qy 6 ProArgLeuGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1307 CCACCCCTGAGAGCCGCCACCCACCCCTGAG-GAGCTCGACCCACCCACCCCTGAG 1365  
Qy 26 ProAspProArgProLeuLeuGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1366 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCACTCCCAAGGAGCGGCT 1422  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1423 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1482  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1483 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1530  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1531 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACTCCC 1578  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1579 AAGAGCTGCTGCCCAAGAGCTTGCACCCACCAACCAAGAGGAGCCACATCCACCAACC 1638  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1639 TCTGACAGCCGCTCCCACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCGCT 1698  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153

Db 1699 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACC 1758  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1759 CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCGCTGCCCAAGAGGCTTGCCACCCACC 1818  
Qy 174 ThrLeuSerThrPro 178  
Db 1819 ACCACCAAGGGGCC 1833

#### RESULT 4

US-07-757-022B-83  
Sequence 83, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/07/757,022B  
APPLICATION NUMBER: 19910910  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3066 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3066  
US-07-757-022B-83

Alignment Scores:  
Pred. No.: 0.0103 Length: 3066  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12

Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-83 (1-3066)

QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
DB 1550 CCACCCCTGAGAACCCGACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1608  
QY 26 ProAspProArgProLeuGluGluGluAlaProLeuLeuProArgThrHisLeu 45  
DB 1609 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGACGGCT 1665  
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpValThrGluProAlaAlaMetThr--- 64  
DB 1666 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1725  
QY 65 -----ProGlyAsnThrProArgThrProGluValThrProLeuArgLeuGlu 82  
DB 1726 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773  
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
DB 1774 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773  
QY 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113  
DB 1822 AAGAAGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCAAGGAGCCACATCCACACC 1881  
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
DB 1882 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1941  
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
DB 1942 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCC 2001  
QY 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
DB 2002 CTCAGGAACCTGCACCACTACTCCCAAGAGCTGCTCCCAAGAGCTGCTCCCAAGAGCT 2061  
QY 174 ThrLeuSerThrPro 178  
DB 2062 ACCACCAAGGGGCC 2076

RESULT 5  
US-07-757-022B-73  
; Sequence 73, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3117 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3114  
US-07-757-022B-73

Alignment Scores:  
Pred. No.: 0.0105 Length: 3117  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-73 (1-3117)

QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
DB 1502 CCACCCCTGAGAACCCGACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1560  
QY 26 ProAspProArgProLeuGluGluGluAlaProLeuLeuProArgThrHisLeu 45  
DB 1561 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGACGGCT 1617  
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpValThrGluProAlaAlaMetThr--- 64  
DB 1618 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1677  
QY 65 -----ProGlyAsnThrProArgThrProGluValThrProLeuArgLeuGlu 82  
DB 1678 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1725  
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
DB 1726 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1773  
QY 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113  
DB 1774 AAGAAGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCAAGGAGCCACATCCACACC 1833  
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
DB 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1893  
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
DB 1894 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCC 1953

Qy 154 ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1954 CTCAAGGAACCTGCACCACTACTCCCAAGAGCTTGCACCCACC 2013  
Qy 174 ThrLeuSerThrPro 178  
Db 2014 ACCACCAAGGGGCC 2028

## RESULT 6

US-07-757-022B-57  
; Sequence 57, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3148 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3147

US-07-757-022B-57  
Alignment Scores:  
Pred. No.: 0.0106 Length: 3148  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-57 (1-3148)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1631 CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1689  
Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGCGGCT 1746  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1747 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1806  
Qy 65 -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1807 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1854  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAenProAspThrGln 102  
Db 1855 -----AAAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1902  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1903 AAGAGCTGCCCCCAAGGAGCTTGCACCCACCACTCCCAAGGAGCCCATCCACCCACC 1962  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1963 TCTGACAAAGCCCGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCT 2022  
Qy 134 GlyCysTrpThrValThrGluProAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2023 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC 2082  
Qy 154 ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2083 CTCAAGGAACCTGCACCACTACTCCCAAGAGCTGCCCCCAAGGAGCTTGCACCCACC 2142  
Qy 174 ThrLeuSerThrPro 178  
Db 2143 ACCACCAAGGGGCC 2157

## RESULT 7

US-07-757-022B-103  
; Sequence 103, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

Db 2356 CTCAAGGAACCTGCACCCACTACTCCCAAGAAGCCTGCCCCCAAGGAGCTTGCACCCACC 2415

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1502 CCACCCCTGAGAACCCCGACCCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1560  
Qy 26 ProAspProArgProLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 45  
Db 1561 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCACTCCCAAGGAGCGGCT 1617  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1618 CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1677  
Qy 65 -----ProGlyAsnThrProArgProArgProArgProGluValThrProLeuArgLeuGlu 82  
Db 1678 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT----- 1725  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1726 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1773  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1774 AAGAGGCTGCTGCCCCAGGAGCTTGCACCCACCCACCCACCCACCCACCCACCCACCC 1833  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCT 1893  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1894 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCC 1953  
Qy 154 ThrGlnGluValThrProLeuLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1954 CTCAAGGAACCTGACCCACCTACTCCCAAGAGGCTGCTCCCAAGAGGCTTGCACCCACCC 2013  
Qy 174 ThrLeuSerThrPro 178  
Db 2014 ACCACCAAGGGGCC 2028

## RESULT 9

US-07-757-022B-41  
; Sequence 41, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3936 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3933  
US-07-757-022B-41  
Alignment Scores:  
Pred. No.: 0.0143 Length: 3936  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-41 (1-3936)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1625 CCACCCCTGAGAACCCCGACCCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1683  
Qy 26 ProAspProArgProLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 45  
Db 1684 ---GAGCCACACCCACCCACCCCTGAGGAGCCTGCTCCACCACTCCCAAGGAGCGGCT 1740  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1741 CCACACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1800  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1801 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCC 1848  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1849 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT----- 1848  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1897 AAGAGGCTGCTGCCCCAGGAGCTTGCACCCACCCACCCACCCACCCACCCACCCACCC 1956  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1957 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCT 2016  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2017 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCC 2076  
Qy 154 ThrGlnGluValThrProLeuLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2077 CTCAAGGAACCTGACCCACCTACTCTCCCAAGAGGCTGCTCCCAAGAGGCTTGCACCCACCC 2136  
Qy 174 ThrLeuSerThrPro 178  
Db 174 ThrLeuSerThrPro 178

Tue Feb 22 09:49:23 2005

```
Db      2137 ACCACCAAGGGGCC 2151
RESULT 10
US-07-757-022B-141
; Sequence 141, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3942 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...3939
US-07-757-022B-141
Alignment Scores:
Pred. No.: 0.0143 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 3 Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-141 (1-3942)
QY      6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db      1631 CCACCCCTGAGAGCCGCCGCCACCCACCCCTGAG-GAGCTGCCACCCACCCCTGAG 1689
26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGCGGGCT 1746
46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
1747 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1806
65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82
1807 AAGGAGCGCTGCTCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 1854
83 LeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAspThrGln 102
1855 -----AAAGGGACTGCTCCAACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1902
103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
1903 AAGAAGCGCTGCCCCCAAGGAGCTTGACCCACCCACCAAGGAGGCCACATCCACACC 1962
114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluHisGlnHis 133
1963 TCTGACAAAGCCCGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACCCCTAAGGAGCCT 2022
134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
2023 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACC 2082
154 ThrGlnGluValThrProLeuLeuGluGlnLysLeuProGluLeuValHisAla 173
2083 CTCGAAGGAACCTGCACCCACTACTCCCAAGAAGCGCTGCCCCCAAGGAGCTTGACCCACC 2142
174 ThrLeuSerThrPro 178
2143 ACCACCAAGGGGCC 2157
RESULT 11
US-07-757-022B-49
; Sequence 49, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3942 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...3939
US-07-757-022B-141
Alignment Scores:
Pred. No.: 0.0143 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 3 Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-141 (1-3942)
QY      6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db      1631 CCACCCCTGAGAGCCGCCGCCACCCACCCCTGAG-GAGCTGCCACCCACCCCTGAG 1689
```



PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3945 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3942  
US-07-757-022B-49

Alignment Scores:  
Pred. No.: 0.0143 Length: 3945  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-49 (1-3945)

Qy	6	ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer	25
Db	1634	CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGACCCACCCCTGAG	1692
Qy	26	ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu	45
Db	1693	---GAGCCACACCCACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGCGGCT	1749
Qy	46	GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr	64
Db	1750	CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT	1809
Qy	65	-----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu	82
Db	1810	AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT-----	1857
Qy	83	LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln	102
Db	1858	-----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGACCCCACTACTCC	1905
Qy	103	AlaSerAlaSerPro-----AspProArgProLeuArg	113
Db	1906	AAGAGCTGCTGCCCCAAGGAGCTTGCACCCACCCACCAAGGAGCCCACTCCACCCACC	1965
Qy	114	GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis	133
Db	1966	TCTGACAGCCGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCT	2025
Qy	134	GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg	153
Db	2026	GCTCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC	2085
Qy	154	ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla	173
Db	2086	CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCCTGCTCCCAAGGAGCTTGCACCCACC	2145
Qy	174	ThrLeuSerThrPro	178
Db	2146	ACCACCAAGGGGCC	2160

RESULT 12

US-07-757-022B-45  
; Sequence 45, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3963 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3960  
US-07-757-022B-45

Alignment Scores:  
Pred. No.: 0.0144 Length: 3963  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-45 (1-3963)

Qy	6	ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer	25
Db	1652	CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGACCCACCCCTGAG	1710
Qy	26	ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu	45

Db 1711 ---GAGCCACACCCACCCAGGAGGCTGCTCCACCACTCCCAAGGACGCGCT 1767  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1768 CCCAACACCCCTAAGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1827  
Qy 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1828 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT----- 1875  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1876 -----AAAGGAGCGCTGCTCCAACTACCCCTCAAGGAACCTGCGACCCCACTACTCCC 1923  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1924 AAGAGCGCTGCTCCCAAGAGGCTTCCACCCACCCACCAAGGAGCGCCACATCCACCAACC 1983  
Qy 114 GluGluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1984 TCTGACAGCGCGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCT 2043  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2044 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACC 2103  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2104 CTCAAGGAACCTGCGACCCCACTACTCCCAAGAGCGCTGCTCCCAAGGAGCGCTGCGACCCCAACC 2163  
Qy 174 ThrLeuSerThrPro 178  
Db 2164 ACCACCAAGGGGCGCC 2178

## RESULT 13

US-07-757-022B-59  
; Sequence 59, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3963 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3960  
US-07-757-022B-59  
Alignment Scores:  
Pred. No.: 0.0144 Length: 3963  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-59 (1-3963)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1652 CCACCCCTGAGAGCGCGCCACCCACCCCTGAG-GAGCTCGCACCCACCCCTCGAG 1710  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1711 ---GAGCCACACCCACCCACCCCTGAGGAGCGCTGCTCCCACTCCCAAGGAGCGCGCT 1767  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1768 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1827  
Qy 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1828 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCCACTACCCCT----- 1875  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1876 -----AAAGGAGCGCTGCTCCAACTACCCCTCAAGGAACCTGCGACCCCACTACTCCC 1923  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1924 AAGAGCGCTGCTCCCAAGAGGCTTCCACCCACCCACCAAGGAGCGCCACATCCACCAACC 1983  
Qy 114 GluGluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1984 TCTGACAGCGCGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCT 2043  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2044 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACC 2103  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2104 CTCAAGGAACCTGCGACCCCACTACTCCCAAGAGCGCTGCTCCCAAGGAGCGCTGCGACCCCAACC 2163  
Qy 174 ThrLeuSerThrPro 178  
Db 2164 ACCACCAAGGGGCGCC 2178  
RESULT 14  
US-07-757-022B-47  
; Sequence 47, Application US/07757022B  
; Patent No. 6433142

GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4065 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..4062  
US-07-757-022B-47  
Alignment Scores:  
Pred. No.: 0.0149 Length: 4065  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-47 (1-4065)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1754 CCACCCCTGAGAGCCGACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1812  
Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1813 ---GAGCCACACCCACCCCTGAGGAGGCTGCTCCACCATCCCAAGGAGCGGCT 1869  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64

Db 1870 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1929  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1930 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT----- 1977  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1978 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGACCCACTACTCCC 2025  
Qy 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113  
Db 2026 AAGAGCGCTGCCCAAGAGCTTGACCCACCCACCAAGGAGCCACATCCACCACC 2085  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 2086 TCTGACAAGCGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCT 2145  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
Db 2146 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCC 2205  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2206 CTCGAAGGAACCTGACCCACTACTCCCAAGGAAGCGCTGCCCAAGGAGCGCTTGACCCACC 2265  
Qy 174 ThrLeuSerThrPro 178  
Db 2266 ACCACCAAGGCGCC 2280  
RESULT 15  
US-07-757-022B-39  
Sequence 39, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822

us-10-030-225-2 copy 27 213.rni

Search completed: February 22, 2005, 03:20:11  
Job time : 464 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 02:21:12 ; Search time 5476 Seconds  
(without alignments)  
201.838 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKKPLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_100357\_2456/app.query.fasta\_1.327  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=inp -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Bloum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10030225 @Cn 1 1 723 @runat 18022005\_100357\_2456  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq\*
- 14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/2/pubna/US10C\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/2/pubna/US10D\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/2/pubna/US10E\_PUBCOMB.seq\*
- 18: /cgn2\_6/ptodata/2/pubna/US10F\_PUBCOMB.seq\*
- 19: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq\*
- 20: /cgn2\_6/ptodata/2/pubna/US11\_NEW\_PUB.seq\*
- 21: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq\*
- 22: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq\*

Pred. No' is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
c 1	658.5	66.2	497	9	US-09-864-761-156 Sequence 156, App
c 2	430	43.2	243	9	US-09-864-761-16984 Sequence 16984, A
c 3	149.5	15.0	6694	17	US-10-221-613-274 Sequence 274, App
	149	15.0	1602	17	US-10-425-114-14755 Sequence 14755, A
c 5	147	15.0	1782	17	US-10-424-599-22752 Sequence 22752, A
c 6	147	14.8	6456	15	US-10-311-455-979 Sequence 979, App
c 7	144	14.5	2824	13	US-10-124-557-13 Sequence 13, Appl
c 8	144	14.5	3066	13	US-10-124-557-83 Sequence 83, Appl
c 9	144	14.5	3117	13	US-10-124-557-73 Sequence 73, Appl
c 10	144	14.5	3148	13	US-10-124-557-57 Sequence 57, Appl
c 11	144	14.5	3420	13	US-10-124-557-103 Sequence 103, App
c 12	144	14.5	3813	13	US-10-124-557-43 Sequence 43, Appl
c 13	144	14.5	3936	13	US-10-124-557-41 Sequence 41, Appl
c 14	144	14.5	3942	13	US-10-124-557-141 Sequence 141, App
c 15	144	14.5	3945	13	US-10-124-557-49 Sequence 49, Appl
c 16	144	14.5	3963	13	US-10-124-557-45 Sequence 45, Appl
c 17	144	14.5	3963	13	US-10-124-557-59 Sequence 59, Appl
c 18	144	14.5	4065	13	US-10-124-557-47 Sequence 47, Appl
c 19	144	14.5	4086	13	US-10-124-557-39 Sequence 39, Appl
c 20	144	14.5	4092	13	US-10-124-557-51 Sequence 51, Appl
c 21	144	14.5	4215	13	US-10-124-557-61 Sequence 61, Appl
c 22	144	14.5	4575	13	US-10-044-090-303 Sequence 303, App
c 23	144	14.5	5008	13	US-10-124-557-1 Sequence 1, Appli
c 24	144	14.5	5041	10	US-09-918-624B-1 Sequence 1, Appli
c 25	144	14.5	5041	11	US-09-897-188-2 Sequence 2, Appli
c 26	144	14.5	5041	17	US-10-641-643-981 Sequence 981, App
c 27	144	14.5	5041	18	US-10-858-595-2 Sequence 2, Appli
c 28	144	14.5	5041	18	US-10-473-974-14 Sequence 14, Appl
c 29	144	14.5	5041	18	US-10-473-974-89 Sequence 89, Appl
c 30	144	14.5	5041	18	US-10-473-974-90 Sequence 90, Appl
c 31	143	14.4	14643	15	US-10-156-761-931 Sequence 931, App
c 32	143	14.4	31422	18	US-10-204-862A-2 Sequence 2, Appli
c 33	143	14.4	9025608	15	US-10-156-761-1 Sequence 1, Appli
c 34	142.5	13.3	88421	9	US-09-976-059-1 Sequence 94182, A
c 35	136.5	13.7	1257	18	US-10-437-963-94182 Sequence 7594, Ap
c 36	136.5	13.7	11294	10	US-09-764-891-7594 Sequence 7595, Ap
c 37	136.5	13.7	15123	10	US-09-764-891-7595 Sequence 8274, A
c 38	136.5	13.7	75853	13	US-10-087-192-382 Sequence 700, App
c 39	135	13.6	2149	18	US-10-437-963-88274 Sequence 1421, Ap
c 40	135	13.6	34722	18	US-10-322-281-700 Sequence 36853, A
c 41	134.5	13.5	2626	17	US-10-108-260A-1421 Sequence 1421, Ap
c 42	133	13.4	885	18	US-10-363-345A-36853 Sequence 36854, A
c 43	133	13.4	885	18	US-10-363-345A-36854 Sequence 3, Appli
c 44	133	13.4	92638	18	US-10-450-826-3 Sequence 47999, A
c 45	132.5	13.3	3725	18	US-10-437-963-47999

ALIGNMENTS

RESULT 1  
US-09-864-761-156/c  
; Sequence 156, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864, 761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03





; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 14755  
 ; LENGTH: 1602  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3039-011-B5\_FLI  
 US-10-425-114-14755

Alignment Scores:  
 Pred. No.: 5,57e-05 Length: 1602  
 Score: 149.00 Matches: 64  
 Percent Similarity: 40.00% Conservative: 20  
 Best Local Similarity: 30.48% Mismatches: 78  
 Query Match: 14.97% Indels: 48  
 DB: 17 Gaps: 8

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-425-114-14755 (1-1602)

Qy 11 ProArgProGlySerLeuThrArgLeuAlaGlu---ValSerAlaSerProAspProArg 29  
 Db 111 CCAGTCCCTCTCTCCAAAGCGGCGGCGATCTCCTCTCTCCACTCTCTCCGTCG 170  
 Qy 30 ProLeuLysGluGluGlu-----GluAlaPro 38  
 Db 171 CCCCTCTCTCTTCCACATCGCTCGCTCTCTCTCGTCGCTCCACGCGCCCT 230  
 Qy 39 LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThr 58  
 Db 231 CTCCTTCTCTCGGACT-----ACTCTGTCT 254  
 Qy 59 GluProAlaAlaMetThrProGlyAsnThrThrProPro-----ArgThr 73  
 Db 255 CTGCTTCGGACCCCTCTCCAGACACATCTGCTCCCTGGAAGCCCTCTCCGCGAG 314  
 Qy 74 ProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThr 93  
 Db 315 CCCACTGGCGCTCTCTACGCGCTCTCTACCGCTCTTCCACCACTGTCTGTCGACAGC 374  
 Qy 94 LeuSer-----ThrProAsnProAspThrGlnAla-----103  
 Db 375 TCAAGCCCCACATCGCCCTCTCCAACTCTCCCTCCCTCCGACTACGCGCTCGCATGG 434  
 Qy 104 SerAla-SerProAspProArgProLeuArgGluGluGluAlaArgLeuLeuProAr 123  
 Db 435 TCCCTCCCGCTCGCCACGCGCTCTCCGCAAAACCTCGCTCCCGCTACTCTCCCTCG 494  
 Qy 123 gThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGluProAlaAl 143  
 Db 495 ACCCTACTCTGCTCTCCAAGATCAACCAATGTCACCGCTCTCTCGCCACCAAGCTCT 554  
 Qy 143 a-----LeuThrProGlyAsnAlaThrProPro-ArgThrGlnGluValThrProLeuL 161  
 Db 555 ACCCTGAGTTATCAAGATCCCGTCGCGCGCGCGCTCTCTCGAAACCAACCCAGGCT 614  
 Qy 161 euLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrPro-----178  
 Db 615 TCGAGGAATCTACCTCGCTCCCAACATGTGGGGCGCCATCGACACCAACCCCGTCCACC 674  
 Qy 179 -----AsnProAspAsnGlnVal 184  
 Db 675 TCCGCAATTAAACCTTAACCTTAACATT 700

RESULT 5

US-10-424-599-22752  
 ; Sequence 22752, Application US/10424599  
 ; Publication No. US20040031072A1

; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 22752  
 ; LENGTH: 1782  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_120547C.1  
 US-10-424-599-22752

Alignment Scores:  
 Pred. No.: 6,25e-05 Length: 1782  
 Score: 149.00 Matches: 64  
 Percent Similarity: 40.00% Conservative: 20  
 Best Local Similarity: 30.48% Mismatches: 78  
 Query Match: 14.97% Indels: 48  
 DB: 17 Gaps: 8

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-424-599-22752 (1-1782)

Qy 11 ProArgProGlySerLeuThrArgLeuAlaGlu---ValSerAlaSerProAspProArg 29  
 Db 138 CCAGTCCCTCTCTCCAAAGCGGCGGCGATCTCCTCTCTCCACTCTCTCCGTCG 197  
 Qy 30 ProLeuLysGluGluGlu-----GluAlaPro 38  
 Db 198 CCCCTCTCTCTTCCACATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 257  
 Qy 39 LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThr 58  
 Db 258 CTCCTTCTCTCGGACT-----ACTCTGTCT 281  
 Qy 59 GluProAlaAlaMetThrProGlyAsnThrThrProPro-----ArgThr 73  
 Db 282 CTGCTTCGGACCCCTCTCCACAGACACATCTGCTCCCTGGAAGCCCTCTCCGCGAG 341  
 Qy 74 ProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThr 93  
 Db 342 CCCACTGGCGCTCTCTTACGCGCTCTCTCTACCCCGCTCTTCCACCACTGTCTGTCGACAGC 401  
 Qy 94 LeuSer-----ThrProAsnProAspThrGlnAla-----103  
 Db 402 TCAAGCCCCACATCGCCCTCTCCAACTCTCTCCCTCCCTCCGACTACGCGCTCGCATGG 461  
 Qy 104 SerAla-SerProAspProArgProLeuArgGluGluGluAlaArgLeuLeuProAr 123  
 Db 462 TCCCTCCCGCTCGCCACGCGCTCTCCGCAAAACCTCGCTCCCGCTACTCTCCCTCG 521  
 Qy 123 gThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGluProAlaAl 143  
 Db 522 ACCCTACTCTGCTCTCCAAGATCAACCAATGTCACCGCTCTCTCGCCACCAAGCTCT 581  
 Qy 143 a-----LeuThrProGlyAsnAlaThrProPro-ArgThrGlnGluValThrProLeuL 161  
 Db 582 ACCCTGAGTTATCAAGATCCCGTCGCGCGCGCGCTCTCTCGAAACCAACCCAGGCT 641  
 Qy 161 euLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrPro-----178  
 Db 642 TCGAGGAATCTACCTCGCTCCCAACATGTGGGGCGCCATCGACACCAACCCCGTCCACC 701  
 Qy 179 -----AsnProAspAsnGlnVal 184  
 Db 702 TCCGCAATTAAACCTTAACCTTAACATT 727



## RESULT 6

US-10-311-455-979/c  
; Sequence 979, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 979  
; LENGTH: 6456  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-979

## Alignment Scores:

Pred. No.:	0.000394	Length:	6456
Score:	147.00	Matches:	57
Percent Similarity:	42.56%	Conservative:	26
Best Local Similarity:	29.23%	Mismatches:	70
Query Match:	14.77%	Indels:	42
DB:	15	Gaps:	11

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-311-455-979 (1-6456)

Qy	11	ProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSerProAspProArgPro 30
Db	2275	CCCCGTCACCTACAAAAACCAAA-----ACCAATCCCCCAACCCCTCCTCC 2228
Qy	31	LeuLysGluGluGluGluAlaProLeu---LeuProArgThrHisLeuGlnAlaGluPro 49
Db	2227	CTCAAAACCGGAAATCCAAACCCCAACCCCTCCTCC-----CTTCAACCCCAAAA 2177
Qy	50	HisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThrProGlyAsnThrThr 69
Db	2176	TCCAAATCTCAACTCTCTCTCTCCCTCCCAACCCCAAAATCCAAACCCCA-----ACCAT 2123
Qy	70	ProProArgThrProGluVal-----ThrProLeuArgLeu 81
Db	2122	CCTCCCTCAAAATCCAAATCCAAACCCCAACCCCTCTCTCTCCCAACCCCAAAATCTA 2063
Qy	82	GluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThr 101
Db	2062	AACCCCAACTCTCTC-----CTCCCTCAAAACCCCAAAATCTAAACCCCAACCCCTCCTCC 2009
Qy	102	GlnAlaSerAlaSerProAspProArgProLeu-----ArgGluGluGluGluAla 118
Db	2008	CTCAAAACCAAAATCCAAACCCCAACCCCTCTCTCTCTCAAAACCCGAAATCCAAACCC 1949
Qy	119	ArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrVal 138
Db	1948	CCAAACCCCTCTCTCTCAAAACCCGAAATCCAA-----ATC 1913
Qy	139	ThrGluProAlaLeuThrPro-----GlyAsnAlaThrProPro----- 152
Db	1912	CCCAACCCCTCTCTCTCACACCTTAAATAATCCAAACCCCAACCCCTCTCTCTCTCAACCCG 1853
Qy	153	-----ArgThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeu 170
Db	1852	AAATTCCAAAACCCCAACCCCTACCTCTCTTCAAAACAAATAATTCAAACCCCAACCCCTCTC 1793

Qy 171 ValHisAlaThrLeuSer-----ThrProAsnProAspAan 182  
Db 1792 CTTCAACCCCAAAATCCAAACCCCAACCCCTCTCTCTCTCAAAC 1748

## RESULT 7

US-10-124-557-13  
; Sequence 13, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: Hewick, Rodney M.  
; APPLICANT: Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989

## ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851

## INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cdna  
FEATURE:

## NAME/KEY: CDS

LOCATION: 1..2823

SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-10-124-557-13

## Alignment Scores:

Pred. No.:	0.000315	Length:	2824
Score:	144.00	Matches:	53
Percent Similarity:	35.14%	Conservative:	12
Best Local Similarity:	28.65%	Mismatches:	99
Query Match:	14.47%	Indels:	22
DB:	13	Gaps:	4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-124-557-13 (1-2824)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25

Db 1307 CCACCCCTGAGAGCCCGCACCACCCCTGAG-GAGCTCGACCCACCCCTTGAG 1365  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuProArgThrHisLeu 45  
Db 1366 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGCGCT 1422  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1423 CCACACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1482  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1483 AAGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1530  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1531 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACTCCC 1578  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1579 AAGAGCTGCTCCCAAGAGCTGTCACCCACCCACCAAGGAGCCCACTCCACCAACC 1638  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1639 TCTGACAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1698  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1699 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 1758  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1759 CTCAGGAACCTGCACCACTACTCCCAAGAGCTGCTCCCAAGAGCTGCTGACCCCAACC 1818  
Qy 174 ThrLeuSerThrPro 178  
Db 1819 ACCACCAAGGGGCC 1833

RESULT 8

US-10-124-557-83  
; Sequence 83, Application US/10124557  
; Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 83:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3066 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3066  
; SEQUENCE DESCRIPTION: SEQ ID NO: 83:  
US-10-124-557-83

Alignment Scores:  
Pred. No.: 0.000344 Length: 3066  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 13 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-124-557-83 (1-3066)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1550 CCACCCCTGAGAGCCCGCACCACCCCTGAG-GAGCTCGACCCACCCCTTGAG 1608  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuProArgThrHisLeu 45  
Db 1609 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCCACTCCCAAGGAGCGCT 1665  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1666 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1725  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1726 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1774 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACTCCC 1821  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1822 AAGAGCTGCTGCCCCCAAGAGCTGTCACCCACCCACCAAGGAGCCCACTCCACCAACC 1881  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1882 TCTGACAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1941  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1942 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 2001  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2002 CTCAGGAACCTGCACCACTACTCCCAAGAGCTGCTCCCAAGAGCTGCTGACCCCAACC 2061  
Qy 174 ThrLeuSerThrPro 178  
Db 2062 ACCACCAAGGGGCC 2076



```

;
; REGISTRATION NUMBER: 31,922
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3148 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3147
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-124-557-57

Alignment Scores:
Pred. No.: 0.000354 Length: 3148
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-57 (1-3148)
QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1631 CCACCCCTGAGAGCCGCCACCCACCCCTCTGAG-GAGCTCGCACCCACCCCTCTGAG 1689
QY 26 ProArgProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1690 ---GAGCCACACCCACCCACCCCTGAGGAGCTGTCTCCACCATCTCCAGGCGAGCGCT 1746
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1747 CCCACACCCCTAAGGAGCGTCTCCAACTACCCCTAAGGAGCGTCTCCAACTACCCCT 1806
QY 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1807 AAGGAGCGTCTCCAACTACCCCTAAGGAGCGTCTCCAACTACCCCT----- 1854
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProArgThrGln 102
Db 1855 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAGCGTCTCCAACTACCCCT 1902
QY 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 1903 AAGAAGCGTGCCTCCCAAGGAGCGTTCGACCCACCCACCCACCCACCATCCACACC 1962
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1963 TCTGACAAAGCGCGTCTCCAACTACCCCTAAGGGAGCGTCTCCAACTACCCCTAAGGAGCGCT 2022
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db 2023 GCTCCAACTACCCCTAAGGAGCGTCTCCAACTACCCCTAAGGGAGCGTCTCCAACTACC 2082
QY 154 ThrGlnGluValThrProLeuLeuLeuGlnLysLeuGlnLysLeuProGluLeuValHisAla 173
Db 2083 CTCAGGAACCTGCACCCCACTACTCCCAAGAAGCGTGTGCCCCCAAGGAGCGTTGCACCCACC 2142
QY 174 ThrLeuSerThrPro 178
Db 2143 ACCACCAAGGGCGCC 2157

RESULT 11
US-10-124-557-103
; Sequence 103, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:

```

```

;
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geertz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3420
; SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-10-124-557-103

Alignment Scores:
Pred. No.: 0.000387 Length: 3420
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-103 (1-3420)
QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1904 CCACCCCTGAGAGCCGCCACCCACCCCTCTGAG-GAGCTCGCACCCACCCCTCTGAG 1962
QY 26 ProArgProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1963 ---GAGCCACACCCACCCACCCCTGAGGAGCGTCTCCACCATCTCCCAAGGCGAGCGCT 2019
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 2020 CCCAACACCCCTAAGGAGCGTCTCCAACTACCCCTAAGGAGCGTCTCCAACTACCCCT 2079

```

Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProGluArgLeuGlu 82  
Db 2080 AAGGAGCCTGCTCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 2127  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAspThrGln 102  
Db 2128 -----AAGGAGCTGCTCAACTACCCCTAAGGAGAACTGACCCACTACTCCC 2175  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 2176 AAGAGCCTGCCCAAGAGCTTGACCCACCACCCAGGAGCCACATCCACCACC 2235  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 2236 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCTAAGGAGCCT 2295  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2296 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACC 2355  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2356 CTCAAGGAGCTGACCCACTACTCCCAAGAGCTGCCCAAGAGCTTGACCCACCACC 2415  
Qy 174 ThrLeuSerThrPro 178  
Db 2416 ACCACCAAGGGGCC 2430

RESULT 12  
US-10-124-557-43  
; Sequence 43, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ceert, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851

; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3813 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3810  
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:  
US-10-124-557-43

Alignment Scores:  
Pred. No.: 0.000436 Length: 3813  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 13 Gaps: 4

US-10-030-225-2\_copy\_27\_213 (1-187) x US-10-124-557-43 (1-3813)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1502 CCACCCTGAGAGCCCGCACCCACCCCTGAG-GAGCTCGCACCCACCCTGAG 1560  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1561 ---GAGCCCAACCCACCCCTGAGGAGCTGCTCCACCCTCCCAAGGAGCGGCT 1617  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1618 CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1677  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1678 AAGGAGCCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1725  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1726 -----AAGGAGCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCTAAGGAGCCT 1773  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1774 AAGAGCCTGCCCAAGGAGCTTGACCCACCACCCAGGAGCCACATCCACCACC 1833  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCTAAGGAGCCT 1893  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1894 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCTAAGGAGCCT 1953  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1954 CTCAGGAGAACTGACCCACTACTCCCAAGAGCCTGCCCAAGAGCTTGACCCACCACC 2013  
Qy 174 ThrLeuSerThrPro 178  
Db 2014 ACCACCAAGGGGCC 2028

## RESULT 13

US-10-124-557-41  
; Sequence 41, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.



```

; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3939
; SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-10-124-557-141

Alignment Scores:
Pred. No.: 0.000452 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-141 (1-3942)

Qy 6' ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1631 CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1689
Qy 26' ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACACTCCCAAGCAGCGGCT 1746
Qy 46' GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1747 CCACACCCCTGAGAGCCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1806
Qy 65' -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1807 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1854
Qy 83' LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1855 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGACCCACTACTCCC 1902
Qy 103' AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 1903 AAGAGCCTGCCCCCAAGGAGCTTGACCCACCCACCCACCCACCATCCACACC 1962
Qy 114' GluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1963 TCTGACAGCCGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCCT 2022
Qy 134' GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAenAlaThrProArg 153
Db 2023 GCTCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC 2082
Qy 154' ThrGlnGluValThrProLeuLeuGluGlnLysLeuProGluLeuValHisAla 173
Db 2083 CTCAAGGAACCTGCACCCACTACTCCCAAGAGCTGCCCCCAAGGAGCTTGACCCACC 2142
Qy 174' ThrLeuSerThrPro 178
Db 2143 ACCACCAAGGGGCC 2157

RESULT 15
US-10-124-557-49
; Sequence 49: Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
```

```

; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3945 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3942
; SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-10-124-557-49

Alignment Scores:
Pred. No.: 0.000452 Length: 3945
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-49 (1-3945)

Qy 6' ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1634 CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1692
Qy 26' ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1693 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACACTCCCAAGCAGCGGCT 1749
Qy 46' GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1750 CCACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1809
Qy 65' -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1810 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1857
Qy 83' LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1858 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGACCCACTACTCCC 1905
Qy 103' AlaSerAlaSerPro-----AspProArgProLeuArg 113
```

Tue Feb 22 09:49:23 2005

```

Db 1906 AAGAAGCTGCCCCCAAGGAGCTTGACCCACCCACCAAGGAGGCCACATCCACCACC 1965
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1966 TCTGACAAGCCCGTCCAACTACCCCTAAGGGGACTGCTCCAACTACCCCTAAGGAGCCT 2025
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db 2026 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACC 2085
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db 2086 CTCAGGAACCTGCACCCACTACTCCCAAGAGCTGCCCCCAAGGAGCTTGCAACCACC 2145
Qy 174 ThrLeuSerThrPro 178
Db 2146 ACCACCAAGGGGCC 2160

```

Search completed: February 22, 2005, 06:10:23  
Job time : 5491 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 01:12:02 ; Search time 3107 Seconds  
(without alignments)  
2290.962 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKPRLRGRPGSLRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_100355\_2354/app\_query.fasta\_1.327  
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -NORM=scorepct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225 @CGN 1 1 5180 @runat\_18022005\_100355\_2354 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hc1:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gse1:\*  
9: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	992	99.7	790	5	BX336289
2	992	99.7	826	5	BX387234
3	992	99.7	866	5	BX365974
4	992	99.7	867	5	BX403263
5	992	99.7	868	5	BX365998
6	992	99.7	872	5	BX344975
7	992	99.7	881	5	BX345009
8	992	99.7	900	5	BX398022
9	992	99.7	917	5	BX334584

10	992	99.7	936	5	BX357282
11	992	99.7	936	5	BX381620
12	992	99.7	938	5	BX397246
13	992	99.7	947	5	BX359737
14	992	99.7	960	5	BX381477
15	992	99.7	967	5	BX357315
16	992	99.7	968	5	BX356306
17	992	99.7	971	5	BX336623
18	992	99.7	973	5	BX379676
19	992	99.7	975	5	BX382082
20	992	99.7	978	5	BX336717
21	992	99.7	979	5	BX335053
22	992	99.7	979	5	BX360184
23	992	99.7	980	5	BX360878
24	992	99.7	982	5	BX359034
25	992	99.7	983	1	AL554329
26	992	99.7	983	5	BX358077
27	992	99.7	983	5	BX361532
28	992	99.7	985	5	BX458619
29	992	99.7	990	5	BX360443
30	992	99.7	994	5	BX339696
31	992	99.7	3009	3	CR749863
32	987	99.2	708	5	BX324767
33	987	99.2	882	5	BX387069
34	987	99.2	924	5	BX378610
35	987	99.2	955	5	BX343795
36	987	99.2	985	3	CNSLT11IBO
37	985	99.0	984	5	BX337459
38	983	98.8	782	5	BX360064
39	982	98.7	960	1	AL549276
40	980	98.5	913	5	BX366004
41	979	98.4	912	5	BX344130
42	978	98.3	870	5	BX387067
43	978	98.3	1113	5	BX439514
44	975	98.0	967	5	BX335752
45	974	97.9	982	5	BX358526

#### ALIGNMENTS

RESULT 1

BX336289

LOCUS

DEFINITION

BX336289 Homo sapiens

clone CS0DI027YA01 5-PRIME, mRNA sequence.

ACCESSION

BX336289

VERSION

BX336289.2

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Li, W.B., Gruber, C.,

Jessie, J. and Polayes, D.

1 (bases 1 to 790)

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 2, 2003 this sequence version replaced gi:30341500.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0DI027AA01Q1P4c=7348.r.

Location/Qualifiers

1. .790

/organism="Homo sapiens"

BX336289 790 bp mRNA linear EST 07-APR-2004  
BX336289 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI027YA01 5-PRIME, mRNA sequence.

ACCESSION  
BX336289

VERSION  
BX336289.2

KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens

REFERENCE  
Li, W.B., Gruber, C.,

Jessie, J. and Polayes, D.

1 (bases 1 to 790)

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 2, 2003 this sequence version replaced gi:30341500.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0DI027AA01Q1P4c=7348.r.

Location/Qualifiers

1. .790

/organism="Homo sapiens"



QY 161 LeuLeuGluInLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 563 CTGTGGAGCTGCAGAAGCTGCCAAGATTGGTCCACGCACTTGAGTAGCCCTAACCGT 622

QY 181 AspAsnGlnValThrIleLys 187  
 Db 623 GATAACCAGGTGACCATCAG 643

```

RESULT 3
BX365974                               866 bp      mRNA       linear     EST 08-APR-2004
DEFINITION   BX365974 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
              clone CSODI031YB04 5-PRIME, mRNA sequence.
ACCESSION    BX365974
VERSION      BX365974.2 GI:46288863
KEYWORDS     Homologous recombination; DNA replication; DNA repair; DNA damage; DNA
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li (bases 1 to 866)
Id,W,B., Gruber,C., Jesses,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30370960.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(QT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7348.r

There is a virtual cDNA representing this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZF12_AX21ZH4_l&c=7348.r
Location/Qualifiers
            1..866
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="CSODI031YB04"
             /tissue_type="PLACENTA COT 25-NORMALIZED"
             /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
             /note="1st strand cDNA was primed with a NotI-oligo (dr)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores:
Pred. No.:          3e-62           Length:          866
Score:             992.00         Matches:        186
Percent Similarity: 100.00%       Conservative:    1
Best Local Similarity: 99.47%     Mismatches:     0
Query Match:       99.70%         Indels:         0
DB:                 5             Gaps:           0

US-10-030-225-2_COPY_27_213 (1-187) x BX365974 (1-866)

QY 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 87 CTCCTCCGTGAAGACGCCGCGCTCCGCGACCAACCGCTGGAGGCCCTCACAGGCTCGCA 146

QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 147 GAGGTCTCAGCCTCCCCAGATCTTAGGCCTCTGAGGNAGNAGGAGGACCACTGCTC 206

QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60

```

	Pred.	No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:	3 922.00 100.00% 99.47% 99.70% 5	Alignment Scores:	Length: Matches: Conservative: Mismatches: Indels: Gaps:	867 186 1 0 0 0
--	-------	--	---	-------------------	---	--------------------------------

US-10-030-225-2 COPY 27 213 (1-187) x BX403263 (1-867)

Qy	1	LeuProValIysGlyProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla	20
Db	106	CTCCCGTGAAGAAGCCGCGGCTCGCGGACACCGGCTGGAGCCTCACGAGGCTCGCA	165
Qy	21	GluValSerAlaSerProAsnProArgProLeuLysGluGluGluAlaProLeuLeu	40
Db	166	GAGGTCTCAGCCTCCCGAGTCTTAGGCTCTGAAGGAGGAGGAGGACCACTGTCTC	225
Qy	41	ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysThrThrValThrGluPro	60
Db	226	CCCGAAGCCACCTCTCAGCGAGAGCCACCAACATGGATGCTGGACTGTCACTGAGCCA	285
Qy	61	AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg	80
Db	286	GCAGCCATACCCCGAGGAACACCAACCCCTCCAGGACCCCGAGAGTTACTCCGTTGCGG	345
Qy	81	LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp	100
Db	346	CTGAGCTCGAAGCTGCGGGATTGGCCAAACAACCTTGAGTACCCCTAACCCCTGAT	405
Qy	101	ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluGluAlaArgLeu	120
Db	406	ACCCAGCCTTCAGCCTCCCGATCTTAGGCCTCTGAGGGAAGAGGAGGACGACTG	465
Qy	121	LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysThrThrValThrGlu	140
Db	466	CTCCCGAAGCAACCACTGTCAGGACAGACTACACCAACTGGATGTTGGACTGTCACTGAG	525
Qy	141	ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu	160
Db	526	CCAGCAGCCCTGACCCCGGGAATGCCAGCCTCCAGGACCCAGGAGGTACTCCCTTG	585
Qy	161	LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro	180
Db	586	CTGCTGGAGCTCGAAGCTGCGAGAATTGGTCCACGCAACCTTGAGTACCCCTAACCT	645
Qy	181	AspAsnGlnValThrIleLys	187
Db	646	GATAACCAAGTGACCATCAAG	666

RESULT 5					
BX365998					
LOCUS	BX365998	868 bp	mRNA	linear	EST 08-APR-2004
DEFINITION	BX365998 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSQDI074Y001 5-PRIME, mRNA sequence.				

ACCESSION BX365998  
VERSION BX365998.2 GI:46306893  
KEYWORDS EST.

KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

ORGANISM	Homio sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE  
1 (bases 1 to 868)

**AUTHORS** Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

**TITLE** Full-length cDNA libraries and normalization

**JOURNAL Unpublished (2001)**

COMMENT On May 5, 2003 this sequence replaced gi:30376966.

**Contact: Genoscope**

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY

Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized, library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348. r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?g=CS2BAX142D10AX242R5.1&c=1>

## FEATURES

Location/Qualifiers  
1. .868

## ORIGIN

Alignment Scores:	3.01e-62	Length:	868
Pred. No.:	992.00	Matches:	186
Score:	100.00%	Conservative:	1
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	99.47%	Indels:	0
Query Match:	99.70%	Gaps:	0
DB:	5		

US-10-030-225-2 COPY 27 213 (1-187) x BX365998 (1-868)

Qy	1	LeuProValLysIleProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla	20
Db	85	CTCCCGTGAAGAAGCCGGCTCCGCGGACACCGGCTGGAGGCTCACGAGGCTCGCA	14
Qy	21	GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu	40
Db	145	GAGGTCCTCAGCCTCCCCAGATCCTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC	20
Qy	41	ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro	60
Db	205	CCAGAACCCACCTGCAGGCAGAGCCACCAACATGGATGCTGACTGTCACTGAGCCA	26
Qy	61	AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg	80
Db	265	GCAGCCATGACCCAGGCAACACCAACCCCTCCAGGACCCAGAGGTACTCCGTGGCG	32
Qy	81	LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp	10
Db	325	CTGAGCTCGAAGCTGCGGGATTGGCCAAACAACCTTGATACCCTAACCTGAT	38
Qy	101	ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluGluAlaArgLeu	12
Db	385	ACCAGGCTTCAGCCTCCCCAGATCCTAGGCCTCTGAGGAAGAGGAGGACGACTG	44
Qy	121	LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu	14
Db	445	CTCCCGAAGCCACCTGCAGGAGAGCTACACCAACATGGATGTGGACTGTCACTGAG	50
Qy	141	ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu	16
Db	505	CCAGAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACCCAGAGGTTACTCCCTTG	56
Qy	161	LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro	18
Db	565	CTGCTGGAGCTGCAGAAGCTGCCAGATTGGTCCACGCAACCTTGAGTACCCCTAACCT	62
Qy	181	AspAsnGlnValThrIleLys	187
Db	625	GATAACCGAGTGCACATCAAG	645

## RESULT 6

BX344975  
 LOCUS BX344975 872 bp mRNA linear EST 08-APR-2004  
 DEFINITION clone CS0D1005YP06 5-PRIME, mRNA sequence.  
 ACCESSION BX344975  
 VERSION BX344975.2 GI:46282437  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 872)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 2, 2003 this sequence version replaced gi:30346327.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r  
 There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS2BAX122A01\_AX19ZH11\_1&c=7348.r

## FEATURES

source

Location/Qualifiers

1. .872  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1005YP06"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3 03e-62 Length: 872  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservativeness: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0  
 US-10-030-225-2\_COPY\_27\_213 (1-187) x BX344975 (1-872)  
 QY 1 LeuProValLysLeuArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 DB 91 CTCCTCGGTGAAGAGCCCGGCTCGCGGACCCAGCCGCTGAGAGCTCAGGAGCTGCA 150  
 QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 DB 151 GAGGTCTCAGGCTCCCGGAGCTCTAGGCTCTGAAGAGAGAGGAGGAGGACCACTGCTC 210  
 QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 DB 211 CCCAGAACCCCTCAGGAGGAGGACCAACCAATGATGCTGACCTGTCTAGGAGCA 270  
 QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 DB 271 GCAGCATGATGACCCAGGCAACACACCCCTCCAGGACCCAGAGGTTACTCCGTGGCG 330  
 QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 DB 331 CTGGAGCTGCAGAGCTCGCGGATTGGCCACACACACACTTGTAGTACCCCTAACCCCTGAT 390

QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 DB 391 ACCCAGGCTTCAGCTCCCGAGATCTCTAGGCTCTGAGGGAAGAGGAGGAGGAGGAGCTG 450  
 QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTTPThrValThrGlu 140  
 DB 451 CTCCTCGGAGAACCCACCTGCGAGGAGAGCTACCAACATGATGTTGACTGTCTACTGAG 510  
 QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 DB 511 CCAGAGCCCTGAGCCCGAGGAGATGCCAGCTCCAGGACCCAGGAGGTTACTCCCTTTG 570  
 QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 DB 571 CTGCTGGAGCTGCAGAGCTGCCAGATTGTCACCGAACCTTGTAGTACCCCTAACCCCT 630  
 QY 181 AspAsnGlnValThrLys 187  
 DB 631 GATAACCCAGGTGACCATCAAG 651

## RESULT 7

BX345009

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact:

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS2BAX14ZE04\_AX232E7\_1&amp;c=7348.r.

Location/Qualifiers

1. .881

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1060YF04"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 3 07e-62 Length: 881

Score: 992.00 Matches: 186

Percent Similarity: 100.00% Conservativeness: 1

Best Local Similarity: 99.47% Mismatches: 0

Query Match: 99.70% Indels: 0

DB: 5 Gaps: 0





Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 917)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 1, 2003 this sequence version replaced gi:30310346.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS0D1004CE11QPI&c=7348.r.  
Location/Qualifiers  
1. .917  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1004YJ21"  
/tissue type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and Ecor V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## FEATURES

source

Alignment Scores:  
Pred. No.: 3.23e-62 Length: 917  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX334584 (1-917)

Qy 1 LeuProVallylsLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 91 CTCCCGGTGAGAGCCGCGCTCCGCGACACCGCTGGAGCCTCACAGGCTCGCA 150  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 151 GAGGTCTCAGGCTCCCGAGATCCTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 210  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 211 CCCAGAACCCACCTGCGAGGAGAGCCACCAACATGGATGCTGACTGTCTACCTGAGCCA 270  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
Db 271 GCAGCCATGATCCCGAGGACCAACACCCCTCCAGGACCCACAGAGTTACTCCGTGGGG 330  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 331 CTGGAGCTGCAGAGCTGCGGGATTGGCCAAACACACACTTGTAGTACCCCTTAACCTGAT 390  
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 391 ACCAGGCTTCAGGCTCCCGAGATCCTAGGCTCTGAGGGAAGAGGAGGAGGACGACTG 450  
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
Db 451 CTCCCGAGAACCCACCTGCGAGGAGAGCTACCAACACATGGATGTTGACTGTCTACTGAG 510  
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160

Db 511 CCACGAGCCCTGACCCCGAGGAATGCCAGCTCCAGGACCCAGGAGTTACTCCCTTG 570  
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 571 CTGCTGAGCTGCAGAGCTGCCAGAAATTGTCACGCAACCTTGAGTACCCCTAACCT 630  
Qy 181 AspAsnGlnValThrLeuLys 187  
Db 631 GATAACCCAGGTGACCATCAAG 651

## RESULT 10

BX357282

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0D1023Y2N3 5-PRIME, mRNA sequence.

BX357282 Homo sapiens PLACENTA COT 25-NORMALIZED

clone CS0D1023Y2N3 5-PRIME, mRNA sequence.

BX357282.1 GI:30384144

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0D1023Y2N3QPI&amp;c=7348.r.

Location/Qualifiers

1. .936

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1023Y2N3"

/tissue type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and Ecor V

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."





1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?8=CS0D1032BE03QPl&c=7348.r>.

## FEATURES

source  
1. .938  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1032YI06"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
Pred. No.: 3.32e-62 Length: 938  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX397246 (1-938)

Qy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 72 CTCCCGGTGAAGACCGCGCTCGCGACACACCGCTGGAGCCTCACGAGCTCGCA 131  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 132 GAGGTCTCAGCCTCCCGAGATCCTAGGCCTCTGAAGGAAGAGGAGGACCACTGCTC 191  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 192 CCCAGAACCCACCTGCAGGACGACCCACCAACATGATGCTGACTGTCTACTGAGCCA 251  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
Db 252 GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCACGAGGTTACTCCGTTGGG 311  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 312 CTGGAGCTGCAGAGCTGCGGGATTGGCCAAACACACCTTGAGTACCCCTAACCTGAT 371  
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 372 ACCAGGCTTCAGCCTCCCGAGATCCTAGGCCTCTGAAGGAAGAGGAGGACGACGTG 431  
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
Db 432 CTCCCGAAGACCCACCTGCAGGACGACCTACCAACATGATGTTGACTGTCTACTGAG 491  
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
Db 492 CCAGCAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACCCAGGAGGTTACTCCCTTG 551  
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 552 CTGCTGGAGCTGCAGAGCTGCCAGAAATTGTTCCACGCAACCTTGAGTACCCCTAACCC 611  
Qy 181 AspAsnGlnValThrIleLys 187  
Db 612 GATAACCAAGGTGACCAATCAAG 632

## RESULT 13

LOCUS : BX359737  
DEFINITION : BX359737 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1060YF04 5-PRIME, mRNA sequence.  
ACCESSION : BX359737  
VERSION : BX359737.2 GI:46305712  
KEYWORDS : EST.  
SOURCE : Homo sapiens (human)  
ORGANISM : Homo sapiens  
REFERENCE :  
1 (bases 1 to 947)  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30376366.  
COMMENT :  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

## FEATURES

source  
1. .947  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1060YF04"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

## ORIGIN

Alignment Scores:  
Pred. No.: 3.36e-62 Length: 947  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX359737 (1-947)

Qy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 92 CTCCCGGTGAAGACCGCGCTCGCGACACACCGCTGGAGCCTCACGAGCTCGCA 151  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 152 GAGGTCTCAGCCTCCCGAGATCCTAGGCCTCTGAAGGAAGAGGAGGACCACTGCTC 211  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 212 CCCAGAACCCACCTGCAGGACGACCCACCAACATGATGCTGACTGTCTACTGAGCCA 271  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
Db 272 GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCAGGAGGTTACTCCGTTGGG 331  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 332 CTGGAGCTGCAGAGCTGCCGGGATTGGCCAAACACACCTTGAGTACCCCTAACCCCTGAT 391

QY 101 ThrGlnAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 |||||  
 Db 392 ACCAGGCTTCAGCTCCAGATCTAGGCTCTGAGGAGAGAGAGAGGACAGCTG 451  
 |||||  
 QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 |||||  
 Db 452 CTCCCAGAACCCACCTGAGGAGAGCTACACCAACATGGATGTTGGACTGCTCACTGAG 511  
 |||||  
 QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 |||||  
 Db 512 CCAGAGGCTTACCCAGGGAATGCCAGCTCCAGGACCCAGGAGGTTACTCCCTTG 571  
 |||||  
 QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 |||||  
 Db 572 CTGTGGAGCTGCAGAGCTGCAGAAATGGTTCACGACCACTTGAGTACCCCTAACCT 631  
 |||||  
 QY 181 AspAsnGlnValThrIleLys 187  
 |||||  
 Db 632 GATAACCAAGTGACCATCAAG 652  
 |||||

RESULT 14  
 BX381477  
 LOCUS  
 DEFINITION BX381477 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0DI066YK17 5-PRIME, mRNA sequence.

ACCESSION  
 VERSION BX381477  
 KEYWORDS EST.

SOURCE  
 ORGANISM Homo sapiens (human)

ORGANISM

REFERENCE  
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 8, 2003 this sequence version replaced gi:30452951.  
 Contact: Genoscope

Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r  
 There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS0DI066AF09QPI&c=7348.r.

FEATURES  
 source

1..960  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI066YK17"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:  
 Pred. No.: 3.41e-62 Length: 960  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX381477 (1-960)

QY 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 |||||  
 Db 92 CTCCCGCTGGAAGACCGCGCTCCGGACCCAGCGCTGGAGCCTCACAGGCTCGCA 151  
 |||||  
 QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 |||||  
 Db 152 GAGGTCTCAGCTCCAGATCTTAGGCTCTGAAGAGAGAGAGGAGGACCACTGCTC 211  
 |||||  
 QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
 |||||  
 Db 212 CCCAGAACCCACCTGCAGGACAGGACACCAACATGGATGCTGACTGTCACTGAGCCA 271  
 |||||  
 QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 |||||  
 Db 272 GCAGCCATGATCCCGAGGCAACACCCCTCCAGGACCCAGAGGTTACTCCGTTCGGG 331  
 |||||  
 QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 |||||  
 Db 332 CTGGAGCTGCAGAGCTGCCGGATTGCCCAACAACCTTGAGTACCCCTAACCTTGAT 391  
 |||||  
 QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 |||||  
 Db 392 ACCAGGCTTCAGGCTCCAGATCTTAGGCTCTGAGGAGAGAGGAGGACGACTG 451  
 |||||  
 QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 |||||  
 Db 452 CTCCCAGAACCCACCTGCAGGACAGCTACCAACATGGATGTTGGACTGCTCACTGAG 511  
 |||||  
 QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 |||||  
 Db 512 CCAGAGGCTTACCCAGGGAATGCCAGCTCCAGGACCCAGGAGGTTACTCCCTTG 571  
 |||||  
 QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 |||||  
 Db 572 CTGTGGAGCTGCAGAGCTGCAGAAATGGTTCACGACCACTTGAGTACCCCTAACCT 631  
 |||||  
 QY 181 AspAsnGlnValThrIleLys 187  
 |||||  
 Db 632 GATAACCAAGTGACCATCAAG 652  
 |||||

RESULT 15  
 BX357315

LOCUS

DEFINITION BX357315 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CS0DI024YG21 5-PRIME, mRNA sequence.

ACCESSION BX357315

VERSION BX357315

KEYWORDS EST.

SOURCE

ORGANISM

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 5, 2003 this sequence version replaced gi:30374142.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0DI024AD11QPI&c=7348.r.

Location/Qualifiers

FEATURES

source 1..967  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS01024YG21"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and Ecor V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

## Alignment Scores:

Pred. No.: 3.45e-62 Length: 967  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX357315 (1-967)

Qy	1	LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla	20
Db	92	CTCCCGGTGAAGAAGCCCGGCTCCGCGGACCAAGGCTGGAGCCTCAGAGGCTCGCA	151
Qy	21	GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu	40
Db	152	GAGGTCTCAGGCTCTCCAGATCTCTAGGCTCTGAGGAGAGGAGGACCACTGCTC	211
Qy	41	ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro	60
Db	212	CCAGACCCACCTGTCAGGACAGAGCCACCAACATGATGCTGACTGTCACTGAGCCA	271
Qy	61	AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg	80
Db	272	GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCAGAGGTTACTCCGTTCGGG	331
Qy	81	LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp	100
Db	332	CTGGAGCTGCARAAGCTCCCGGATTCGCCAACACACCTTGATACCCCTAACCCCTGAT	391
Qy	101	ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu	120
Db	392	ACCCAGGCTTCAGCTCCCGAGATCCTAGGCTCTGAGGAGAGGAGGAGGACGACTG	451
Qy	121	LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu	140
Db	452	CTCCCGAGACCCACCTGTCAGGACAGCTACCAACATGGATGTTGACTGTCACTGAG	511
Qy	141	ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu	160
Db	512	CCAGAGCCCTGACCCAGGGAATGCCACGCTCCAGAGCCAGGAGGTTACTCCCTTG	571
Qy	161	LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro	180
Db	572	CTGCTGGAGCTGCAGAGCTGCCAGATTGGTCCACGCAACCTTGAGTACCCCTAACCCCT	631
Qy	181	AspAsnGlnValThrIleLys	187
Db	632	GATAACCAAGTGACCATCAAG	652

Search completed: February 22, 2005, 03:12:19  
Job time : 3111 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 20:55:10 ; Search time 2878 Seconds  
(without alignments)  
10758.472 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696  
Perfect score: 639  
Sequence: 1 atgcgtgcgtccgcgaccg.....ataaccaggtgaccatcaag 639

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.Ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	639	100.0	2981	6	AX136231 Sequence
2	639	100.0	2981	6	BD093285 Sequence
3	639	100.0	2981	6	BD123568 Secretory
4	639	100.0	2981	9	AK075445 Homo sapi
5	635.8	99.5	2883	9	AJ583024 Homo sapi
6	542.6	84.9	750	6	AX136531 Sequence
7	542.6	84.9	750	6	BD123771 Secretory
8	522.4	81.8	3052	9	AK056709 Homo sapi
9	497.4	77.8	1698	6	C0720015 Sequence
10	475.2	74.4	2783	9	BC033140 Homo sapi
11	370	57.9	231464	9	AF111168 Homo sapi
12	261.8	41.0	497	6	CQ049341 Sequence
13	261.8	41.0	497	6	CQ064365 Sequence
14	261.8	41.0	497	6	CQ091307 Sequence
15	261.8	41.0	497	6	C0130141 Sequence
16	261.8	41.0	497	6	CQ168762 Sequence
17	261.8	41.0	497	6	CQ197880 Sequence
18	261.8	41.0	497	6	CQ213333 Sequence
19	261.8	41.0	497	6	CQ251912 Sequence

C 20	261.8	41.0	497	6	CQ289061 Sequence
C 21	261.8	41.0	497	6	CQ326064 Sequence
C 22	239.8	37.5	243	6	CQ054407 Sequence
C 23	239.8	37.5	243	6	CQ073647 Sequence
C 24	239.8	37.5	243	6	CQ104532 Sequence
C 25	239.8	37.5	243	6	CQ143244 Sequence
C 26	239.8	37.5	243	6	CQ178734 Sequence
C 27	239.8	37.5	243	6	CQ203077 Sequence
C 28	239.8	37.5	243	6	CQ226434 Sequence
C 29	239.8	37.5	243	6	CQ264575 Sequence
C 30	239.8	37.5	243	6	CQ301666 Sequence
C 31	239.8	37.5	243	6	CQ338879 Sequence
C 32	143	22.4	801	9	HS4330362 Homo sapi
C 33	125.2	19.6	172662	10	AC120540 Mus muscu
C 34	55.2	8.6	125020	9	AF429315 Homo sapi
C 35	51	8.0	125020	9	AF429315 Homo sapi
C 36	46.4	7.3	11687	1	AE004662 Pseudomon
C 37	44.8	7.0	2704	6	CQ843353 Sequence
C 38	44.8	7.0	2704	9	AK126288 Homo sapi
C 39	44.8	7.0	112027	9	AC073502 Homo sapi
C 40	44.4	6.9	110000	1	BX571965 33 o
C 41	44.4	6.9	110000	1	CP000010_24 Continuation (34 o
C 42	43.2	6.8	1116	6	BD180389 Highly th
C 43	43.2	6.8	273285	1	AE017304 Thermus t
C 44	43	6.7	3177	9	AF055989 Homo sapi
C 45	43	6.7	110477	4	CR456637 Platypus

ALIGNMENTS

RESULT 1  
AX136231  
LOCUS AX136231 2981 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 153 from Patent EP1067182.  
ACCESSION AX136231  
VERSION AX136231.1 GI:14272639  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and  
Hayashi, K.  
TITLE Secretory protein or membrane protein  
JOURNAL Patent: Ep 1067182-A 153 10-JAN-2001;  
Helix Research Institute (JP)  
FEATURES  
source  
1. .2981  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
58. .1773  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39758.1"  
/db\_xref="GI:14272640"  
/translation="MRALDRAGLLLCVLLALLLEALGLPVKKPRLRPPGSLTR  
LAEVASPDPRPLKEEEELPRTTHQAEPHQGWTVTEPAAMTPGNTTPPTPEV  
TPRLLEQLPLGLASTLSTPNDTQASDPRLPREEEELRLPRTTHQAELHQH  
CAWVTEPALTPGNATPRTQEVTPPLLEQLKPLVHATLTPNPNQVTKVVEDP  
QNVSPIDDLAEPNPPQDTLSMLPALMPFLWDGDKGEKDRAPGEKEEEDVYP  
SEIDEGEDQDEDEEEQALWNGTTDNDQGLARGDVFKDSVSYDYPEQKWS  
WSPCGNCTGKQRTKPCGCTATRTCTDLPSCGTEDKDTGLGPSWKLGLARN  
ATDMHQDVDSCKWLNKSDFLIKYLSQMLRDLPCPCAYFLAMPDPSVSLQDEHQ  
RSFRWRDASGFRRLDIYPTARFLRSLSGESSTLAQHCYDEDESLTRTRKGAG  
MPLNLSTDFSPKLHFKPDTTPWILCKGWSRLHAVLPNNNGRACTDNPBEEYLQ  
EAXEY"

ORIGIN

Query Match 100.0%; Score 639; DB 6; Length 2981;  
Best Local Similarity 100.0%; Pred. No. 6.5e-129;

Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGGCTCCGGACCGAGCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 60  
DB 58 ATGGGTGGCTCCGGACCGAGCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 117  
QY 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCGGAGACCAACCGCTGGG 120  
DB 118 CTGAGGCGGGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCGGAGACCAACCGCTGGG 177  
QY 121 AGCCTCAGAGGCTGCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 180  
DB 178 AGCCTCAGAGGCTGCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 237  
QY 181 GAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 240  
DB 238 GAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 297  
QY 241 TGGACTGTCACTGAGCGAGCGCATGACCCAGGCAACACACCCCTCCCGAGACCCCA 300  
DB 298 TGGACTGTCACTGAGCGAGCGCATGACCCAGGCAACACACCCCTCCCGAGACCCCA 357  
QY 301 GAGGTACTCCGTTGCGGCTGAGAGTGCAGAGCTGCGGAGTGGCGAGACCAACCTTG 360  
DB 358 GAGGTACTCCGTTGCGGCTGAGAGTGCAGAGCTGCGGAGTGGCGAGACCAACCTTG 417  
QY 361 AGTACCCCTAACCTGATACCGAGGCTTCCAGGCTCCCGAGATCTCAGGCTCTGAGGAA 420  
DB 418 AGTACCCCTAACCTGATACCGAGGCTTCCAGGCTCCCGAGATCTCAGGCTCTGAGGAA 477  
QY 421 GAGGAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAGGAA 480  
DB 478 GAGGAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAGGAA 537  
QY 481 TGTGGACTGTCACTGAGCGAGCGAGCTGACCCAGGAGTGGCGAGTGGCGAGACCAACCTTG 540  
DB 538 TGTGGACTGTCACTGAGCGAGCGAGCTGACCCAGGAGTGGCGAGTGGCGAGACCAACCTTG 597  
QY 541 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTCGTCAGGCTCCAGCAACC 600  
DB 598 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTCGTCAGGCTCCAGCAACC 657  
QY 601 TTGAGTACCCCTAACCTGATACCAAGGTGACCATCAAG 639  
DB 658 TTGAGTACCCCTAACCTGATACCAAGGTGACCATCAAG 696

RESULT 2  
LOCUS BD093285 2981 bp DNA linear PAT 27-AUG-2002  
DEFINITION Differentiation growth factor.  
ACCESSION BD093285  
VERSION BD093285.1 GI:22638873  
KEYWORDS WO 0104312-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Yoshida, K. and Masuho, Y.  
TITLE Differentiation growth factor  
JOURNAL Patent: WO 0104312-A 1 18-JAN-2001;  
HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,  
YURI KAWAI, KENJI YOSHIDA, YASUHIKO MASUHO  
OS Homo sapiens (human)  
PN WO 0104312-A/1  
PD 18-JUN-2001  
PF 06-JUL-2000 WO 2000JP004514  
PR 08-JUL-1999 JP 99P 194179, 18-OCT-1999 US 60/159586 PI  
TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, KENJI PI  
YOSHIDA,  
PI YASUHIKO MASUHO  
PC C12N15/16, C12N15/12, C12N15/85, C12N5/10, C12P21/02, C07K14/575,

PC C07K14/72,  
PC C12Q1/68, C12Q1/02, A61K67/027//C12P21/02, C12R1/91 CC  
FH Key Location/Qualifiers  
FT CDS (58)..(1770).  
FEATURES  
source Location/Qualifiers  
1..2981  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 639; DB 6; Length 2981;  
Best Local Similarity 100.0%; Pred. No. 6.5e-129;  
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGTGGCTCCGGACCGAGCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 60  
DB 58 ATGGGTGGCTCCGGACCGAGCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 117  
QY 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCCGAGACCAACCGCTGGG 120  
DB 118 CTGAGGCGGGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCCGAGACCAACCGCTGGG 177  
QY 121 AGCCTCAGAGGCTGCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 180  
DB 178 AGCCTCAGAGGCTGCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 237  
QY 181 GAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 240  
DB 238 GAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAAGAAGAG 297  
QY 241 TGGACTGTCACTGAGCGAGCGCATGACCCAGGCAACACACCCCTCCCGAGACCCCA 300  
DB 298 TGGACTGTCACTGAGCGAGCGCATGACCCAGGCAACACACCCCTCCCGAGACCCCA 357  
QY 301 GAGGTACTCCGTTGCGGCTGAGAGTGCAGAGCTGCGGAGTGGCGAGTGGCGAGACCAACCTTG 360  
DB 358 GAGGTACTCCGTTGCGGCTGAGAGTGCAGAGCTGCGGAGTGGCGAGTGGCGAGACCAACCTTG 417  
QY 361 AGTACCCCTAACCTGATACCGAGGCTTCCAGGCTCCCGAGATCTCAGGCTCTGAGGAA 420  
DB 418 AGTACCCCTAACCTGATACCGAGGCTTCCAGGCTCCCGAGATCTCAGGCTCTGAGGAA 477  
QY 421 GAGGAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAGGAA 480  
DB 478 GAGGAGGAGGACCACTGCTCCCGAGAGGTCTCAGCCTCCCGAGATCTCAGGCTCTGAGGAA 537  
QY 481 TGTGGACTGTCACTGAGCGAGCGAGCTGACCCAGGAGTGGCGAGTGGCGAGACCAACCTTG 540  
DB 538 TGTGGACTGTCACTGAGCGAGCGAGCTGACCCAGGAGTGGCGAGTGGCGAGACCAACCTTG 597  
QY 541 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTCGTCAGGCTCCAGCAACC 600  
DB 598 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTCGTCAGGCTCCAGCAACC 657  
QY 601 TTGAGTACCCCTAACCTGATACCAAGGTGACCATCAAG 639  
DB 658 TTGAGTACCCCTAACCTGATACCAAGGTGACCATCAAG 696

RESULT 3  
LOCUS BD123568 2981 bp DNA linear PAT 18-SEP-2002  
DEFINITION Secretory protein or membrane protein.  
ACCESSION BD123568  
VERSION BD123568.1 GI:23218513  
KEYWORDS JP 2002017376-A/77.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and









```

Qy 479 GATGTTGACTGTCACTGAGCAGCAGCCCTGAGCCACCCAGGAATGCCACGCTCC-AGG 537
Db 538 GATGTTGACTGTCACTGAGCAGCAGCCCTGAGCCACCCAGGAATGCCACGCTCC-AGG 597
Qy 538 ACCAGGAGGTTACTCCCTTCTGCTGAGCTGCA--GAACTGCGAG-AAATGTTGTCAC 594
Db 598 ACCAAGANGTTACTCCCTTCTGCTGAGCTGCAAGAAAGCTGCCAGAAATGTTGTCAC 657
Qy 595 GCAA--CCTTGAGTACCCCTAACCTGATACACAGG 628
Db 658 GCAAACCTTGAGTTANCCCTAACCTGATACACAG 693

RESULT 8
AK056709 3052 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.
DEFINITION AK056709
ACCESSION AK056709
VERSION 1 Gi:16552189
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Ohayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,Y., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,A., Sudo,H., Hosiiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamiyama,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yanazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togai,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kunagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Naguchi,S., Itoh,T., Shigetake,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2
Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
Isogai,T.
NEO human cDNA sequencing project
Unpublished
3 (bases 1 to 3052)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

```

```

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES             Location/Qualifiers
     source            1. 3052
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="PLACE5000116"
                        /tissue_type="placenta"
                        /clone_lib="PLACE5"
                        /note="cloning vector: pME18SFL3"

ORIGIN
Query Match      81.8%; Score 522.4; DB 9; Length 3052;
Best Local Similarity 85.8%; Pred. No. 1.6e-103;
Matches 638; Conservative 0; Mismatches 1; Indels 105; Gaps 1;

Qy 1 ATGCGTGGCTCCCGCAGCCGAGCGGGCTCTCTCTCTGCGTGTCTGCTGGCGCGCTG 60
Db 24 ATGCGTGGCTCCCGCAGCCGAGCGGGCTCTCTCTCTGCGTGTCTGCTGGCGCGCTG 83
Qy 61 CTGAGAGCGCGCTGAGGGCTCCCGTGAAGAGCGCGGCTCCCGGACACACGCTGGG 120
Db 84 CTGAGAGCGCGCTGAGGGCTCCCGTGAAGAGCGCGGCTCCCGGACACACGCTGGG 143
Qy 121 AGCTTCACGAGGCTCGAG----- 139
Db 144 AGCTTCACGAGGCTCGAGAGCGCGCTTTTCAGGTGCGCGCTAGCTCGGAGAGGTCCC 203
Qy 140 ----- 139
Db 204 GCGGTGCGAGGATCTGCGGCTTTTGGAGTGGCGGCGCGCGCTGAGGTTCAAAC 263
Qy 140 ----AGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGGAGGACCACTG 195
Db 264 GGTGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGGAGGACCACTG 323
Qy 196 CTCCCGAGAGCCACCTGAGGAGAGCCACCAACATGATGCTGACTGTCACTGAG 255
Db 324 CTCCCGAGAGCCACCTGAGGAGAGCCACCAACATGATGCTGACTGTCACTGAG 383
Qy 256 CCAGCAGCCATGACCCCGAGCAACACACCCCTCCCGAGGACCCCGAGGTTACTCCGTTG 315
Db 384 CCAGCAGCCATGACCCCGAGCAACACACCCCTCCCGAGGACCCCGAGGTTACTCCGTTG 443
Qy 316 CGGTGAGCTGAGAGCTGCGGGATGCGGAGAGCAACCTTGTAGTACCCCTAACCT 375
Db 444 CGGTGAGCTGAGAGAGCTGCGGGATGCGGAGAGCAACCTTGTAGTACCCCTAACCT 503
Qy 376 GATACCCAGGCTTCAGGCTCCCGAGATCTTAGGCTCTGAGGAGAGAGGAGGACGACGA 435
Db 504 GATACCCAGGCTTCAGGCTCCCGAGATCTTAGGCTCTGAGGAGAGAGGAGGAGGACGACGA 563
Qy 436 CTGCTCCCGAGAACCCACCTGCGAGGAGAGCTACCAACATGATGTTGACTGTCACT 495
Db 564 CTGCTCCCGAGAACCCACCTGCGAGGAGAGCTACCAACATGATGTTGACTGTCACT 623
Qy 496 GAGCAGCAGCCCTGAGCCCGAGGAGAGTCCAGGCTCCCGAGGACCCAGGAGGTTACTCCC 555
Db 624 GAGCAGCAGCCCTGAGCCCGAGGAGAGTCCAGGCTCCCGAGGACCCAGGAGGTTACTCCC 683
Qy 556 TTGCTGCTGGAGCTGCAGAGAGCTGCCAGAAATTGGTCCACGCAACTTGTAGTACCCCTAAC 615
Db 684 TTGCTGCTGGAGCTGCAGAGAGCTGCCAGAAATTGGTCCACGCAACTTGTAGTACCCCTAAC 743

```

QY 616 CCTGATACCAAGGTGACCATCAAG 639  
 Db 744 CCTGATACCAAGGTGACCATCAAG 767

RESULT 9  
 CQ720015

LOCUS CQ720015 1698 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 5949 from Patent WO02068579.  
 ACCESSION CQ720015  
 VERSION CQ720015.1 GI:42280872  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
 TITLE Kits, such as nucleic acid arrays, comprising a majority of  
 human exons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 5949 06-SEP-2002;  
 PE Corporation (NY) (US)  
 FEATURES source  
 Location/Qualifiers  
 1..1698  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 77.8%; Score 497.4; DB 6; Length 1698;  
 Best Local Similarity 99.8%; Pred. No. 4.6e-98;  
 Matches 498; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 141 GGTCTCAGCTCCCGCAGATCTTAGGCTCTGAAGAGAGAGAGGAGGACACACTGCTCC 200  
 Db 402 GGTCTCAGCTCCCGCAGATCTTAGGCTCTGAAGAGAGAGAGGAGGACACACTGCTCC 461

QY 201 CAGAACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 260  
 Db 462 CAGAACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 521

QY 261 AGCATGACCCAGGCAACACACCCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 320  
 Db 522 AGCATGACCCAGGCAACACACCCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 581

QY 321 GGAGCTCAGAGCTGGCGGATGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 380  
 Db 582 GGAGCTCAGAGCTGGCGGATGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 641

QY 381 CCAGGCTTCAGCTCCCGCAGATCTTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 440  
 Db 642 CCAGGCTTCAGCTCCCGCAGATCTTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 701

QY 441 CCCAGAGACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 500  
 Db 702 CCCAGAGACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 761

QY 501 AGCAGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 560  
 Db 762 AGCAGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 821

QY 561 GCTGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 620  
 Db 822 GCTGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 881

QY 621 TAACAGG 639  
 Db 882 TAACAGG 900

RESULT 10  
 BC033140

LOCUS BC033140 2783 bp mRNA linear PRI 19-JUL-2004  
 DEFINITION Homo sapiens thrombospondin, type I, domain containing 3, mRNA  
 (CDNA clone IMAGE:3844836), partial cds.  
 ACCESSION BC033140  
 VERSION BC033140.1 GI:21619879  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2783)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Aitschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Retchem, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smaluk, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 2783)  
 Strausberg, R.  
 Direct Submission  
 Submitted (25-JUN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgaps-x@mail.nih.gov](mailto:cgaps-x@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,  
 Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth  
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,  
 Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,  
 Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lissa Prabhu,  
 Parvaneh Saeedi, JR Santos, Angelique Scherch, Ursula Skalek,  
 Duane Smaluk, Jeff Stott, Miranda Tsai, George Yang, Jacquie  
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 43 Row: c Column: 10.  
 Features Location/Qualifiers  
 1..2783  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3844836"  
 /tissue type="Placenta, choriocarcinoma"  
 /clone\_lib="NIH MGC 21"  
 /lab\_host="DH10B-R"  
 /notes="Vector: pOTB7"  
 <1..2783  
 /gene="THSD3"



EST contigs that includes AA397580, A1138314, AA399545, AA68184, all from testis. The closest match in BLASTX is to E. coli pyroline-5-carboxylate reductase."

/codon\_start=1  
/product="unknown"  
/protein\_id="AAD09625.1"  
/db\_xref="GI:186186"  
/translation="MDMLQDLSELOFYGVPEEDRIWLYQGRSRGLMIRACAHATFF  
CKLYNRASLNKQSRHLSIGLSNATPEEFKVGIGGHGKQLAGTLLOLGPTP  
AESLRITRPERTEGELKIGKCFYHNADLVSWADVIFLCCLPSQPNICVEIYTSL  
EKASIVSVFAAIPLRLKLLNHTNLRPOYQYDEDSVSWGANKGVIAALQDPITL  
QATCPSPAGHGSYFVHSENYFPVGLISY"  
complement(9631..9708)

/rpt\_family="L1NE2"

/rpt\_family="L1NE2"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="AluX"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

/rpt\_family="L143\_Send"

Query Match

Best Local Similarity

Matches 494; Conservative

0; Mismatches 5; Indels 111; Gaps 1;

Qy 133

Db 87523

Qy 193

Db 87463

Qy 253

Db 87403

Qy 313

Db 87343

Qy 373

Db 87283

Qy 382

Db 87223

Qy 382

Db 87163

Qy 382

Db 87104

Qy 382

Db 87104

Qy 382

Db 87104



Search completed: February 21, 2005, 23:10:33  
Job time : 2884 secs



**This Page Blank (uspto)**